R´esum´e du Cours de Statistique Descriptive

Yves Till´e18 janvier 2008

Ob jectif et moyens

Objectifs du cours – Apprendre les principales techniques de statistique descriptive univari´ee et bivari´ee.

– ˆEtre capable de mettre en oeuvre ces techniques de mani`ere appropri´ee dans un contexte donn´e
– ˆEtrecapable d'utiliserles commandes debase duLanguage R. Pouvoir appliquer les techniques de statistiques descriptivesau moyen du language R.
– R´ef´erencesDo dge Y.(2003), *Premiers pas en statistique*, Springer.

Dro esb eke J.-J. (1997), *´El´ements de statistique*, Editions de l'Universit´elibre de Bruxelles/Ellipses.

Moyens – 2heuresde coursparsemaine.– 2heuresde TPparsemaine, r´epartisenTP th´eoriquesetapplicati

Le language R – Shareware : gratuit et install´e en 10 minutes.– Open source (on sait ce qui est r´e

Table des matières

Chapitre1

Variables, données statistiques, tableaux, effectifs

1.1 D´efinitions fondamentales

1.1.1 La science statistique

– M´ethodescientific

– Etymologiquement: science de l'´etat.– La statistique s'applique `a la plupart des disciplines : agronom
biologie,démographie,économie,
sociologie,linguistique,psychologie,...

1.1.2 Mesure et variable

– On s'int´eresse `a des *unit´es statis*

des en-
treprises, des m´enages. En sciences humaines,on s'int´eresse dans la plupart des cas `a un nombrefi
d'unit´es.– Sur ces unit´es, on mesure un caract`ere ou une *variable* , le chiffre d'affaires de l'entreprise,
le revenu
du m´enage, l'ˆage de la p ersonne, la cat´egorie so cio-professionnelle d'une personne. On suppose q
variable prend toujours une seule valeur sur chaque unit´e. Les variables sont d´esign´ees par simplici
par une lettre (X,Y,Z).– Les *valeurs possibles* de la variable, sont appel´ees *modalit´es*.– L'ensemble des val
fait de pouvoir ou non ordonner les modalit´es est parfois discutable. Par exemple :dans les cat´egories soc
Une variable est dite continue, si l'ensemble des valeurs possibles estcontinu.**Remarque 1.1** Ces d´efinitio

Exemple 1.1 Les modalités de la variable *sexe* sont *masculin* (codé M) et *féminin* (codé F). Le domaine de la variable est $\{M , F\}$.

Exemple 1.2 Les modalités de la variable nombre d'enfants par famille sont 0,1,2,3,4,5,... C'est une variable quantitative discrète.

1.1.4 Série statistique

On appelle *série statistique* la suite des valeurs prises par une variable X sur n unités d'observation.
Le nombre d'unités d'observation est n.
Les valeurs de la variable X sont notées

$$x_1, ..., x_i, ..., x_n.$$

Exemple 1.3 On s'intéresse à la variable "état-civil" notée X et à la série statistique des valeurs prises par X sur 20 personnes. La codification est C: célibataire, M : marié(e), V : veuf(ve), D : divorcée. Le domaine de

$M\ M\ D\ C\ C\ M\ C\ C\ C\ M C\ M\ V\ M\ V\ D\ C\ C\ C\ M$ Ici, n=20, $x_1 = M, x_2 = M, x_3 = D, x_4 = C, x_5 = C, ..., x_{20} = M$.

1.2 Var

l'effectif de la modalité x_j. La fréquence d'une modalité est l'effectif divisé par le nombre d'unités d'obs

En langageR

```
>X=c('Mari´e(e)','Mari´e(e)','Divorc´e(e)','C´elibataire','C´elibataire','Mari´e(e)','C´elibataire',
      'C´elibataire','C´elibataire','Mari´e(e)','C´elibataire','Mari´e(e)','Veuf(ve)','Mari´e(e)',
      'Veuf(ve)','Divorc´e(e)','C´elibataire','C´elibataire','C´elibataire','Mari´e(e)')
>T1=table(X)> V1=c(T1)> data.frame(Eff=V1,Freq=V1/sum(V1))
```

Eff FreqC´elibataire 9 0.45Divorc´e(e) 2 0.10Mari´e(e) 7 0.35Veuf(ve) 2 0.10**1.2.2 Diagrammeensect**

Le tableau statistique peut ˆetre repr´esent´e par un diagramme en barres ou en secteurs (ou camemb
ou *piechart*en anglais) (voirFigures 1.1 et 1.2). Célibataire

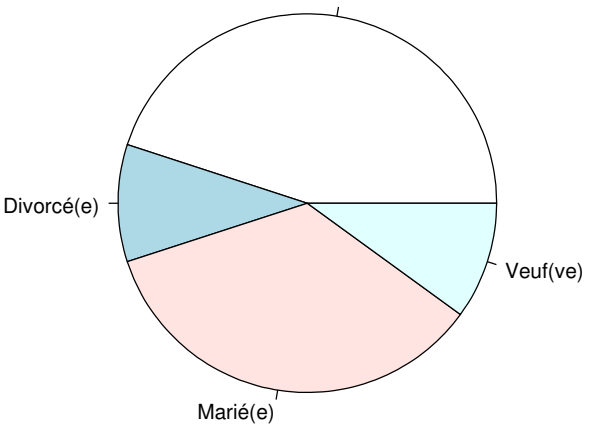


Fig. 1.1 – Diagramme en secteurs**En langage R**> pie(T1,radius=1.0)**En langage R**> barplot(T1)**1.3 Va**

Fig.1.2–Diagramme enbarres

La notation $x_1 \prec x_2$ se lit x_1 pr´ec`ede x_2 .Si la variable est ordinale, on peut calculer les effectifs cumul´es:

$$N_j = \sum_{k=1}^{j} n_k.$$

On a $N_1 = n_1$ et $N_J = n$. On peut ´egalement calculer les fr´equences cumul´ees

$F_j = \frac{N_j}{n} = \sum_{k=1}^{j} f_k$.**Exemple 1.5** On interroge 50 personnes sur leur dernier dipl^ome obtenu (variable Y).

La co dificationa
´et´efaite selon le Tableau 1.1. On a obtenu la s´erie statistique pr´esent´ee dans le tableau 1.2. Finalemen
obtientletableaustatistiquecompletpr´esent´edansleTableau1.3.Tab.1.1–Codificationdelavariable$Y$Dernierdip
U**EnlangageR**>YY=c("Sd","Sd","Sd","Sd","P","P","P","P","P","P","P","P","P","P","P","Se","Se","Se","Se","Se","

Tab. 1.3– Tableau statistique complet

x_j	n_j	N_j	f_j	F_j
Sd	4	4	0.08	0.08
P	11	15	0.22	0.30
Se	14	29	0.28	0.58
Su	9	38	0.18	0.76
U	12	50	0.24	1.00
	50		1.00	

T2=table(YF)V2=c(T2)> data.frame(Eff=V2,EffCum=cumsum(V2),Freq=V2/sum(V2),FreqCum=cumsum(V2/

Eff EffCum Freq FreqCumSd 4 4 0.08 0.08P 11 15 0.22 0.30Se 14 29 0.28 0.58Su 9 38 0.18 0.76U 12 50 0.24

Figure 1.3). Sd

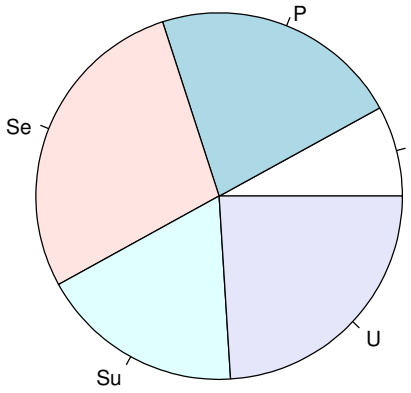


Fig.1.3– Diagrammeen secteursdesfr´equences**En langage R**> pie(T2,radius=1)9

1.3.3 Diagramme enbarres deseffectifs

Les effectifsd'une variablequalitative sont repr´esent´esau moyend'un diagramme en barres (voir Figure 1.4). Sd

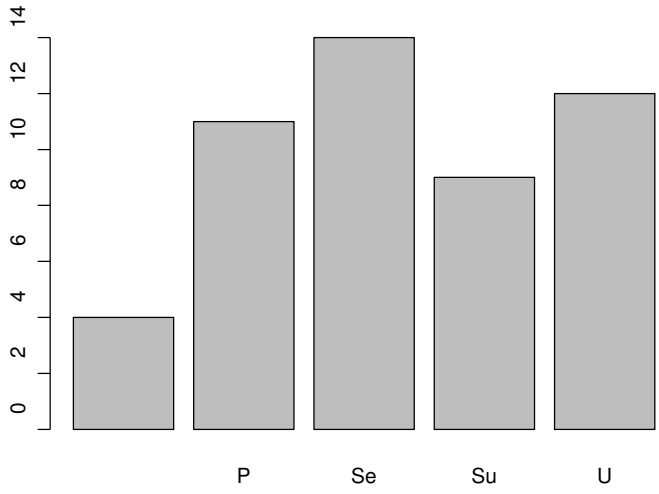


Fig. 1.4– Diagrammeenbarres deseffectifs

En langage R> barplot(T2)**1.3.4 Diagramme en barres des effectifs cumul´es**Leseffectifs

Figure 1.5). Sd

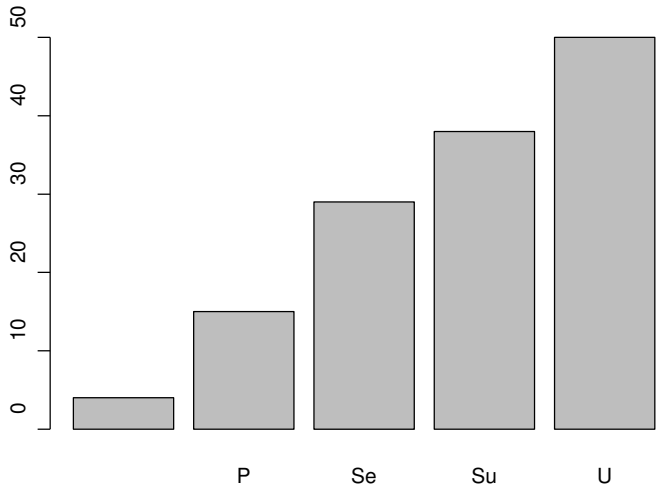


Fig. 1.5–Diagramme en barresdes effectifs cumul´es10

En langageR

```
>T3=cumsum(T2)
>barplot(T3)
```

1.4 Variable quantitative discrète

1.4.1 Le tableaustatistique

Une variable discr`etea un domaine de d´efinitionfini ou d´enombrable.

Exemple 1.6 Un quartier est compos´e de 50 m´enages, et la variable Z repr´esente le nombre de personnes par m´enage.Les valeursde lavariablesont1 1 1 11

					2	2	2	2	2
2	2	2	2	3	3	3	3	3	3
3	3	3	3	3	3			3	3	4
4	4	4	4	4	4	4	4			4	5
5	5	5	5	5	6	6	8				8

Comme p our les variables qualitatives ordinales, on peut calculer les effectifs, les effectifs cumul´es, les fr´equences, les fr´equences cumul´ees. `A nouveau, on peut construire le tableau statistique:

x_j	n_j	N_j	f_j	F_j
1	5	5	0.10	0.10
2	9	14	0.18	0.28
3	15	29	0.30	0.58
4	10	39	0.20	0.78
5	6	45	0.12	0.90
6	3	48	0.06	0.96
8	2	50	0.04	1.00

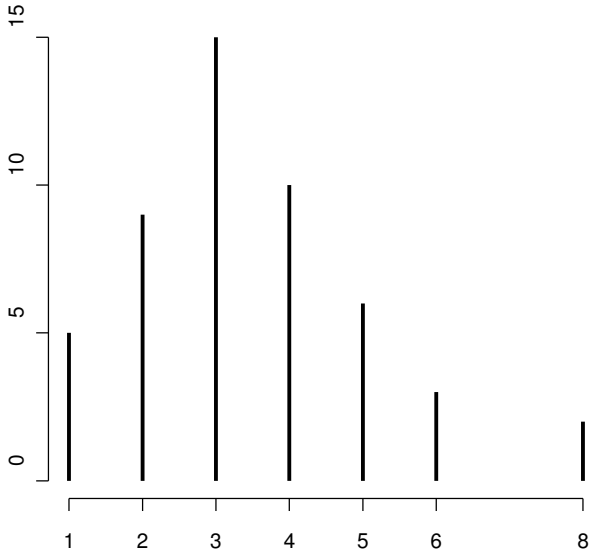


Fig. 1.6 – Diagramme en bˆatonnets des effectifs pour une variable quantitative discr`ete

1.4.2 Diagramme en bˆatonnets des effectifs

Quand la variableest discr`ete, les effectifs sont

En langage R> plot(T4,type="h",xlab="",ylab="",main="",frame=0,lwd=3)

1.4.3 Fonction de r´epa

esent´e

enFigure1.7,est d´efiniedeR dans[0,1]etvaut:$F(x)=\begin{cases} 0 & x<x_1 \\ F_j & x_j \le x<x_{j+1} \\ 1 & x_J \le x. \end{cases}$ **En langage R**> plot(ecdf(
discr`etes. Cependant, il est souvent int´eressant de proc´eder `a des regroupements en classes pourfaire

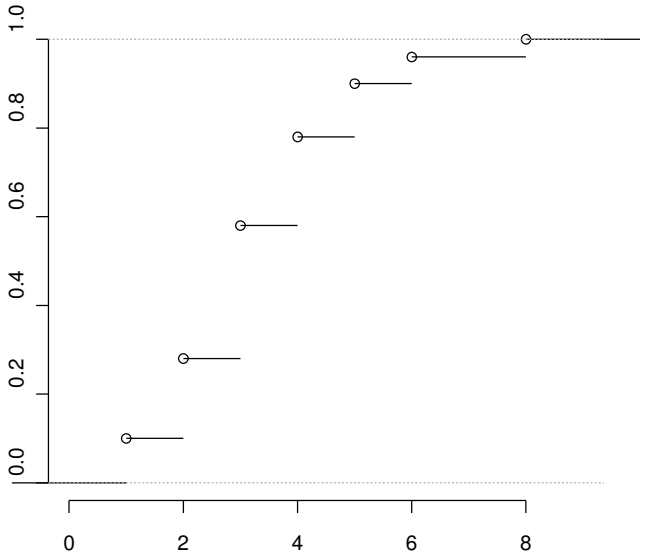


Fig. 1.7 – Fonction de r´epartition d’une variable quantitative discr`ete

Exemple 1.7 On mesure la taille de 50 ´el`eves d’une classe :152 152 152 153 153154 154 154 155 1551
50 0.20 1.0050 1.00Le tableau regroup´e en classe est souvent appel´e *distribution group´ee*. On note, de

– N_j l'effectif cumulé de la classe j,
– f_j la fréquence de la classe j,
– F_j la fréquence cumulée de la classe j.

En langage R> S=c(152,152,152,153,153,154,154,154,155,155,156,156,156,156,156,

+157,157,157,158,158,159,159,160,160,160,161,160,160,161,162,
+162,162,163,164,164,164,164,165,166,167,168,168,168,169,169,
+ 170,171,171,171,171)> T5=table(cut(S, breaks=c(151,155,159,163,167,171)))

>T5c=c(T5)> data.frame(Eff=T5c,EffCum=cumsum(T5c),Freq=T5c/sum(T5c),FreqCum=cumsum(T5c/sum(T

Eff EffCum Freq FreqCum(151,155] 10 10 0.20 0.20(155,159] 12 22 0.24 0.44(159,163] 11 33 0.22 0.66(163

hauteur) représente l'effectif. La hauteur h_j du rectangle correspondant à la classe j est donc donnée par

$$h_j = n_j c_{+j} - c_{-j}.0$$

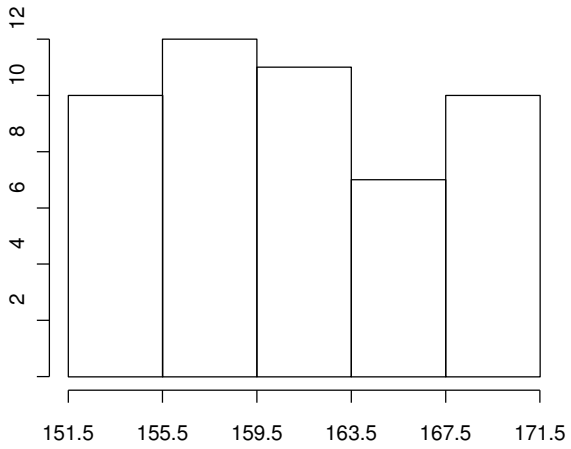


Fig.1.8–Histogrammedeseffectifs**En langage R**> hist(S,breaks=c(151.5,155.5,159.5,163.5,167.5,171.5),xla

Si les deux derni`eres classes sont regroup´ees, comme dans la Figure 1.9, la surface du dernier rectangle est ´egale `a la surface des deux derniers rectangles de l'histogramme de la Figure 1.8.

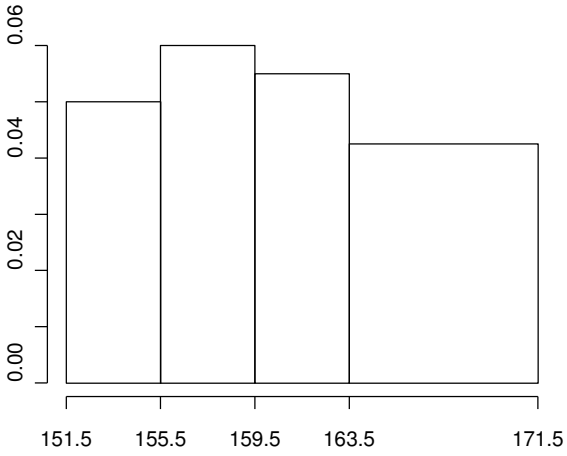
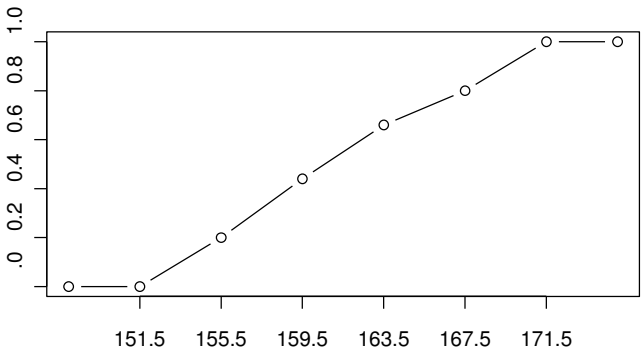


Fig. 1.9– Histogramme des effectifs avec les deux derni`eres classes regroup´ees

En langage R> hist(S,breaks=c(151.5,155.5,159.5,163.5,171.5),xlab="",ylab="",main="",xaxt = "n")> ax



En langageR

```
>y=c(0,0,cumsum(T5c/sum(T5c)),1)
>x=c(148,151.5,155.5,159.5,163.5,167.5,171.5,175)
> plot(x,y,type="b",xlab="",ylab="",xaxt ="n")
> axis(1, c(151.5,155.5,159.5,163.5,167.5,171.5))
```

Chapitre2

Statistique descriptive univari´ee

2.1 Param`etres de position

2.1.1 Le mode

Le mode est la valeur distincte correspondant `a l'effectif le plus ´elev´e; il est not´e x_M.

Si on reprend la variable 'Etat civil' , dontle tableau statistique est le suivant:

x_j	n_j	f_j
C	9	0.45
M	7	0.35
V	2	0.10
D	2	0.10
	n= 20	1

le mode est C: c´elibataire.**Remarque 2.1**– Le mode peut ˆ

2.1.2 La moyenne

La *moyenne* ne peut ˆetre d´efinie que sur une

On peut aussi faire les calculs avec les valeurs distinctes et les effectifs en le tableau: On le tableau

x_j	n_j
0	2
1	3
2	1
3	1
4	1
	8

$$\bar{x} = \frac{2\times 0 + 3\times 1 + 1\times 2 + 1\times 3 + 1\times 4}{8}$$
$$= \frac{3 + 2 + 3 + 4}{8}$$

= 1.5.**Remarque 2.2** La moyenne n'est pas n´ecessairementune valeurpossible.

En langage RE=c(0,0,1,1,1,2,3,4)n=length(E)xb=sum(E)/nxbxb=mean(E)xb**2.1.3 Remarques sur l**
$x_{31} + x_{32}$ $(i = 3)$18

5. On peut exclure une valeur de l'indice.

$$\sum_{\substack{i=1\\ i=3}}^{5} x_i = x_1 + x_2 + x_4 + x_5.$$

Propri´et´e 2.1 1. Somme d'une constante $\sum_{i=1}^{n} a = a + a + \cdots + a$

$$\underbrace{\qquad}_{n \text{ fois}} = na \qquad \text{(a constante)}.$$

Exemple $\sum_{i=1}^{5} 3 = 3 + 3 + 3 + 3 + 3 = 5 \times 3 = 15$.

2. Mise en ´evidence $\sum_{i=1}^{n} a x_i = a \sum_{i=1}^{n} x_i$ (a constante).

Exemple $\sum_{i=1}^{3} 2 \times i = 2(1 + 2 + 3) = 2 \times 6 = 12$.3. Somme des n premiers entiers $\sum_{i=1}^{n} i = 1 + 2 + 3 + \cdots + n = n($

2.1.4 Moyenne géométrique

Si $x_i \geq 0$, on appelle moyenne géométrique la quantité

$$G = \left(\prod_{i=1}^{n} x_i \right)^{1/n} = (x_1 \times x_2 \times \cdots \times x_n)^{1/n} = \exp \frac{1}{n} \sum_{i=1}^{n} \log x_i .$$

La moyenne géométrique s'utilise, par exemple, quand on veut calculer la moyenne de taux d'intérêt.

Exemple 2.3 Supposons que les taux d'intérêt pour 4 années consécutives soient respectivement de 5, 10, 15, et 10%. Que va-t-on obtenir après 4 ans si je place 100 francs?

– Après 1 an on a, $100 \times 1.05 = 105$ Fr.
– Après 2 ans on a, $100 \times 1.05 \times 1.1 = 115.5$ Fr.
– Après 3 ans on a, $100 \times 1.05 \times 1.1 \times 1.15 = 132.825$ Fr.
– Après 4 ans on a, $100 \times 1.05 \times 1.1 \times 1.15 \times 1.1 = 146.1075$ Fr.

Si on calcule la moyenne arithmétique des taux on obtient

$$\bar{x} = \frac{1.05 + 1.10 + 1.15 + 1.10}{4} = 1.10.$$

Si on calcule la moyenne géométrique des taux, on obtient $G = (1.05 \times 1.10 \times 1.15 \times 1.10)^{1/4}$

$$= 1.099431377.$$

Le bon taux moyen est bien G et non $\bar{x}$, car si on applique 4 fois le taux moyen G aux 100 francs, on obtient

$$100 \text{ Fr} \times G^4 = 100 \times 1.099431377^4 = 146.1075 \text{ Fr}.$$

2.1.5 Moyenne harmonique Si $x_i \geq 0$, on appelle moyenne harmonique la quantité $H = n / \sum_{i=1}^{n} 1/x_i$. Il e
$20h50 = 20.8333h$, sa vitesse moyenne est donc Moy$= 400/20.8333 = 19.2$ km/h. – Si on calcule la moyenne

Remarque 2.3 Il est possible de montrer que la moyenne harmonique est toujours inférieure ou égale `a la moyenneg´eom´etrique qui est toujours inférieure ou égale `a la moyenne arithmétique

$$H \leq G \leq \bar{x}.$$

2.1.6 Moyenne pond´er´ee

Danscertains cas, on n'accorde par le mˆeme poids `a toutes les observations. Par exemple, si on calcule la moyenne des notes pour un programme de l'étudiant, on peut pondérer les notes de l'étudiant par le nombre de cr´editsou parle nombred'heures dechaquecours. Si $w_i > 0$, $i = 1, .., n$ sont les poids associés à chaque observation, alors la moyenne pond´er´ee par w_i est définie par :

$$\bar{x}_w = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}.$$

Exemple 2.5 Supp osons que les notes soient pond´er´ees par le nombre de crédits, c'est-à-dire que les notes de l'´etudiant soient les suivantes : Note 5 4 3 6 5Cr´edits 6 3 4 3 4La moyenne pond´er´ee des notes par les cr´edits est donc

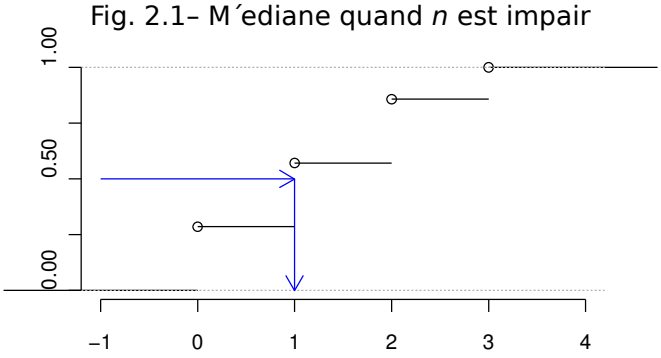
$$\frac{5 \times 6 + 4 \times 3 + 3 \times 4 + 6 \times 3 + 5 \times 4}{6 + 3 + 4 + 3 + 4} = \frac{92}{20} = 4.6.$$

2.1.7 La m´ediane

La m´ediane, not´ee $x_{1/2}$, est une valeur centrale de la s´erie statistique obtenue d

– On trie la s´erie statistique par ordre croissant des valeurs observ´ees. Avec la s´erie observ´ee :3 2 1 ´epartition pour la valeur 1/2 : $x_{1/2} = F^{-1}(0.5)$. **En langage R** 21

Fig. 2.1– M´ediane quand *n* est impair



x=c(0 , 0 , 1 , 1 , 2 ,2 , 3)median(x)plot(ecdf(x),xlab="",ylab="",main="",frame=FALSE,yaxt = "n")

axis(2, c(0.0,0.25,0.50,0.75,1.00))arrows(-1,0.5,1,0.50,length=0.14,col="blue")

arrows(1,0.50,1,0,length=0.14,col="blue")– Si *n* estpair,deuxvaleurssetrouventau milieudela s´

erie (ici avec *n*= 8)

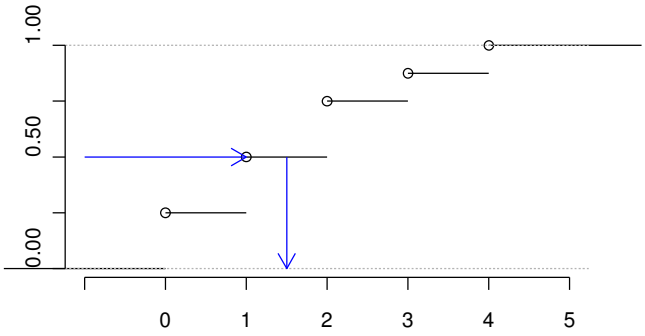
0 01 1 2 234

↑ ↑La m´ediane est alors la moyenne de ces deux valeurs :$x_{1/2}$ = 1+22 = 1.5.La Figure 2.2 montre la

ediane peut toujours

ˆetre d´efinie comme l'inverse de la fonction de r´epartition p our la valeur 1/2 :$x_{1/2} = F^{-1}(0.5)$.Cep er

exactement`aun 'palier'. Fig. 2.2– M´edianequand *n* estpair–1



Enlangage Rx=c(0 , 0 , 1 , 1 , 2 , 2 , 3 , 4)median(x)plot(ecdf(x),xlab="",ylab="",main="",frame=FALSE,ya

En général on note

$$x_{(1)}, \ldots, x_{(i)}, \ldots, x_{(n)}$$

la série ordonnée par ordre croissant. On appelle cette série ordonnée la statistique d'ordre. Cette notation, très usuelle en statistique, permet de définir la médiane de manière très synthétique.

– Si n est impair

$$x_{1/2} = x_{(\frac{n+1}{2})}$$

– Si n est pair

$$x_{1/2} = \frac{1}{2}\left(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}\right).$$

Remarque 2.4 La médiane peut être calculée sur des variables quantitatives et sur des variables qualitatives ordinales.

2.1.8 Quantiles

La notion de quantile d'ordre p (où $0 < p < 1$) généralise la médiane. Formellement un quantile est donné par l'inverse de la fonction de répartition : $x_p = F^{-1}(p)$.

Si la fonction de répartition était continue et strictement croissante, la définition du quantile serait sans équivoque. La fonction de répartition est cependant discontinue par palier. Quand la fonction de répartition est par palier, il existe au moins 9 manières différentes de définir des quantiles selon que l'on fasse ou non une interpolation de la fonction de répartition. Nous présentons une de ces méthodes. Il ne faut pas s'étonner de voir les valeurs des quantiles différer légèrement d'un logiciel statistique à l'autre.

– Si np est un nombre entier, alors $x_p = \frac{1}{2}\left(x_{(np)} + x_{(np+1)}\right)$. – Si np n'est pas un nombre entier, alors $x_p = x_{(\lfloor np \rfloor + 1)}$. Soit la série statistique 12, 13, 15, 16, 18, 19, 22, 24, 25, 27, 28, 34 contenant 12 observations ($n = 12$). – Le premier qua

– Le troisième quartile : Comme $np = 0.75 \times 12 = 9$ est un nombre entier, ona

$$x_{3/4} = \frac{x_{(9)} + x_{(10)}}{2} = \frac{25 + 27}{2} = 26.$$

En langage Rx=c(12,13,15,16,18,19,22,24,25,27,28,34)

quantile(x,type=2)**Exemple 2.7** Soitlas´eriestatistique 12,13, 15, 16,18, 19, 22, 24,25, 27 contenant10obs

– Le premier quartile : Comme $np = 0.25 \times 10 = 2.5$ n'est pas un nombre entier, ona

$$x_{1/4} = x_{(2.5)} = x_{(3)} = 15.$$

– La m´ediane : Comme $np = 0.5 \times 10 = 5$ est un nombre entier, ona

$$x_{1/2} = \underline{12}\, x_{(5)} + x_{(6)} = (18 + 19)/2 = 18.5.$$

– Le troisi`eme quartile : Comme $np = 0.75 \times 10 = 7.5$ n'est pas un nombre entier, ona

$x_{3/4} = x_{(7.5)} = x_{(8)} = 24$.**En langage R**x=c(12,13,15,16,18,19,22,24,25,27)quantile(x,type=2)**2.2 Para**

2.2.3 La variance

La *variance* est la somme des carrés des écarts à la moyenne divisé par le nombre d'observations:

$$s_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Théorème 2.1 *La variance peut aussi s'écrire*

$$s_x^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2. \qquad (2.1)$$

Démonstration $s_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 =$

$$\frac{1}{n}\sum_{i=1}^{n}(x_i^2 - 2x_i\bar{x} + \bar{x}^2)$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - 2\frac{1}{n}\sum_{i=1}^{n} x_i\bar{x} + \frac{1}{n}\sum_{i=1}^{n}\bar{x}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - 2\bar{x}\frac{1}{n}\sum_{i=1}^{n} x_i + \bar{x}^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - 2\bar{x}\bar{x} + \bar{x}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2.$$

✷

La variance peut également être définie à partir des effectifs et des valeurs distinctes:

$s_x^2 = \frac{1}{n}\sum_{j=1}^{J} n_j(x_j - \bar{x})^2$. La variance peut aussi s'écrire $s_x^2 = \frac{1}{n}\sum_{j=1}^{J} n_j x_j^2 - \bar{x}^2$. Quand on veut estimer une va

s'électionnée au hasard) de taille n, on utilise la variance "corrigée" divisée par $n-1$. $S_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i -$
$2+3+4+4+5+6+7+98 = 5{,}25$

$$s_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$= \frac{1}{8}(2-5)^2 + (3-5)^2 + (4-5)^2 + (4-5)^2 + (5-5)^2 + (6-5)^2 + (7-5)^2 + (9-5)^2$$

$$= \frac{1}{8}[9 + 4 + 1 + 1 + 0 + 1 + 4 + 16]$$

$= \frac{36}{8} = 4.5$. On peut ´egalement utiliser la formule (2.1) de la variance,

$$s_x^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2$$

ce qui n´ecessite moins de calculs (surtout quand la moyenne n'est pas un nombre entier).

$$s_x^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2$$

$$= \frac{1}{8}(2^2 + 3^2 + 4^2 + 4^2 + 5^2 + 6^2 + 7^2 + 9^2) - 5^2$$

$$= \frac{1}{8}(4 + 9 + 16 + 16 + 25 + 36 + 49 + 81) - 25$$

$= \frac{236}{8} - 25 = 29.5 - 25 = 4.5$. **Enlangage R**

```
> x=c(2,3,4,4,5,6,7,9)
> n=length(x)
> s2=sum((x-mean(x))^2)/n
```

2.2.5 L'écart moyen absolu

L'*écart moyen absolu* est la somme des valeurs absolues des écarts à la moyenne divisée par le nombre d'observations:

$$e_{moy} = \frac{1}{n}\sum_{i=1}^{n} |x_i - \bar{x}|.$$

2.2.6 L'écart médian absolu

L'*écart médian absolu* est la somme des valeurs absolues des écarts à la médiane divisée par le nombre d'observations :

$$e_{med} = \frac{1}{n}\sum_{i=1}^{n} x_i - x_{1/2}.$$

2.3 Moments

Définition 2.2 *On appelle moment à l'origine d'ordre* $r \in \mathbb{N}$ *le paramètre*

$$m_r = \frac{1}{n}\sum_{i=1}^{n} x_i^r.$$

Définition 2.3 *On appel le moment centré d'ordre* $r \in \mathbb{N}$ *le paramètre*

$m_r = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^r$ .Les moments généralisent la plupart des paramètres. On a en particulier– $m_1 = \bar{x}$

etrie

et l'aplatissement.

2.4 Paramètres de forme

2.4.1 Coefficient d'asymétrie de F

normalisée par la distance interquartile : $A_Y = x_{3/4} + x_{1/4} - 2x_{1/2}x_{3/4} - x_{1/4}$ .27

2.4.3 Coefficient d'asymétrie de Pearson

Le coefficient d'asymétrie de Pearson est basé sur une comparaison de la moyenne et du mode, et est standardisé par l'écart-type:

$$A_P = \frac{\bar{x} - x_M}{s_x}.$$

Tous les coefficients d'asymétrie ont les mêmes propriétés, ils sont nuls si la distribution est symétrique, négatifs si la distribution est allongée à gauche (left asymmetry), et positifs si la distribution est allongée à droite (right asymmetry) comme montré dans la Figure 2.3.

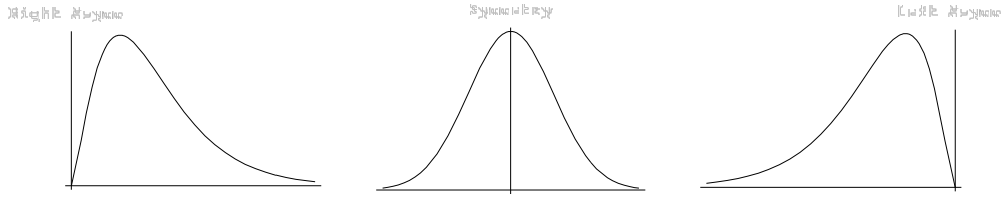


Fig. 2.3– Asymétrie d'une distribution

Remarque 2.6 Certaines variables sont toujours très asymétriques à droite, comme les revenus, les tailles des entreprises, ou des communes. Une méthode simple pour rendre une variable symétrique consiste alors à prendre le logarithme de cette variable.

2.5 Paramètre d'aplatissement (kurtosis)

courtes. Dans la Figure 2.4, on présente un exemple de deux distributions de même moyenne et de même

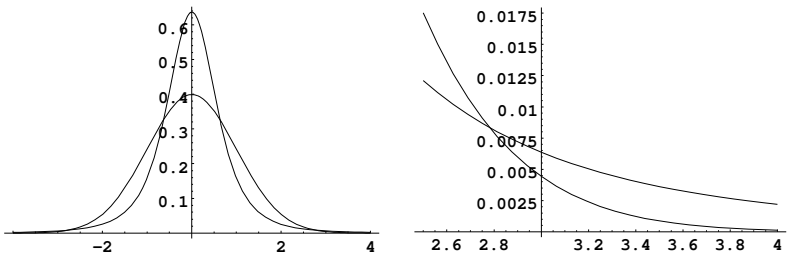


Fig. 2.4– Distributions mésokurtique et leptokurtique

2.6 Changementd’origine et d’unit´

D´efinition 2.4 *On appel le changement d’origine l’op´eration consistant `a ajouter (ou soustraire) la m^eme quantit´e a∈* R *`a toutes les observations*

$$y_i = a + x_i,\ i = 1, \ldots, n$$

D´efinition 2.5 *On appel le changement d’unit´e l’op´eration consistant `a multiplier (ou diviser) par la m^eme quantit´e b∈* R *`a toutes les observations*

$$y_i = bx_i,\ i = 1, \ldots, n.$$

D´efinition 2.6 *On appel le changement d’origine et d’unit´e l’op´eration consistant `a multiplier toutes les observations par la m^eme quantit´e b∈* R *puis `a ajouter la m^eme quantit´e a∈* R *`a toutes les observations:*

$$y_i = a + bx_i,\ i = 1, \ldots, n.$$

Th´eor`eme 2.2 *Si oneffectueun changementd’origine et d’unit´e sur une variable x, alors sa moyenne est affect´ee du m^eme changement d’origine et d’unit´e.* **D´emonstration** Si $y_i = a + bx_i$, alors $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} (a$

$$\bar{} \qquad \bar{} \sum_{i=1} x_i = a + b\bar{x}.$$

✷

Th´eor`eme 2.3 *Si on effectueun changement d’origine et d’unit´e sur une variable x, alors sa varianceest affect´ee par le carr´e du changement d’unit´eet pas parle changement d’origine.*

D´emonstration Si $y_i = a + bx_i$, alors $s^2_y = \frac{1}{n}\sum_{i=1}^{n} (y_i - \bar{y})^2 = \frac{1}{n}\sum_{i=1}^{n} (a + bx_i - a - b\bar{x})^2 = b^2 \frac{1}{n}\sum_{i=1}^{n} (x_i -$

$$\bar{} \qquad \bar{} \qquad \bar{} \qquad s_x^2.$$

✷

Remarque 2.7 1. Les param`etres de position sont tous affect´es par un changement d’origine et d’unit´e.

d’origine.3. Les param`etres de forme et d’aplatissement ne sont affect´es ni par un changement d’unit´e n

$, x_{n_A+1}, x_{n_A+2}, \ldots, x_{n-1}, x_n$ observations de G_B .On d´efinit les moyennes des deux groupes : 29

– la moyenne du premier groupe est $\bar{x}_A = \frac{1}{n_A} \sum_{i=1}^{n_A} x_i$,

– la moyenne du deuxième groupe est $\bar{x}_B = \frac{1}{n_B} \sum_{i=n_A+1}^{n} x_i$.

La moyenne générale est une moyenne pondérée par la taille des groupes des moyennes des deux groupes.
En effet $\bar{x} =$

$$\frac{1}{n} \left(\sum_{i=1}^{n_A} x_i + \sum_{i=n_A+1}^{n} x_i \right) = \frac{1}{n} (n_A \bar{x}_A + n_B \bar{x}_B).$$

On peut également définir les variances des deux groupes:

– la variance du premier groupe est $s_A^2 = \frac{1}{n_A} \sum_{i=1}^{n_A} (x_i - \bar{x}_A)^2$,

– la variance du deuxième groupe est $s_B^2 = \frac{1}{n_B} \sum_{i=n_A+1}^{n} (x_i - \bar{x}_B)^2$.

Théorème 2.4 *(de Huygens) La variance totale est définie par*

$$s_x^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2,$$

se décompose de la manière suivante : $s_x^2 = n_A s_A^2 + n_B s_B^2$ *variance intra-groupes* $+ n_A(\bar{x}_A - \bar{x})^2 + n_B(\bar{x}_B - \bar{x})^2$ *variance inter-groupes.*

Démonstration $s_x^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^{n_A} (x_i - \bar{x})^2 + \sum_{i=n_A+1}^{n} (x_i - \bar{x})^2$ (2.2)

On note que $\sum_{i=1}^{n_A} (x_i - \bar{x})^2 = \sum_{i=1}^{n_A} (x_i - \bar{x}_A + \bar{x}_A - \bar{x})^2 = \sum_{i=1}^{n_A} (x_i - \bar{x}_A)^2 + \sum_{i=1}^{n_A} (\bar{x}_A - \bar{x})^2 + 2 \sum_{i=1}^{n_A} (x_i - \bar{x}_A)(\bar{x}_A$

2.8 Diagramme en tigesetfeuilles

Lediagrammeentigeset feuillesou *Stem and leaf diagram*estune manière rapide de présenter une variable quantitative. Par exemple, si l'on a la statistique ordonnée suivante:

15,15, 16, 17, 18, 20,21, 22, 23,23, 23, 24, 25, 25, 26,

26,27, 28, 28, 29, 30,30, 32, 34,35, 36, 39, 40, 43, 44,

la tige du diagramme sera les dizaines et les feuilles seront les unités. On obtient le graphique suivant.

The decimal point is 1 digit(s) to the right of the|

1 | 556782 | 0123334556678893 | 00245694 | 034Ce diagrammepermet d'avoirune vuesynthétiquede lad

Évidemment,les tiges peuvent être d´efinies par les centaines, ou des millers, selon l'ordre de grandeur de la variable étudiée.

En langage R## Diagramme en tige et feuilles#X=c(15,15,16,17,18,20,21,22,23,23,23,24,25,25,26,26,27

qui permet de repr´esenter la distribution d'une variable. Ce diagramme est comp os´e de :– Un rectangle

pondant `a la m´ediane.– Ce rectangle est compl´et´e par deux segments de droites.– Pour les dessiner, on extrˆemes".**Exemple 2.9** On utilise une base de donn´ees de communes suisses de 2003 fournie par l'Offic ´etriedeladistribution,ausensoùilyabeaucoupdepetitescommunesetpeudegrandescommunes.Legraphiquen

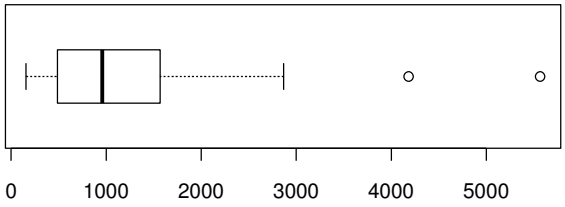


Fig. 2.5 – Boˆıtes `a moustaches p our la variable sup erficie en hectares (HApoly) des communes du canton de Neuchˆatel# ´Etape 1: installation du package sampling

dans lequel se trouve la basede donn´eesdes communes belges
choisir "sampling"dansla liste
utils:::menuInstallPkgs()# Etape 2: charge le package sampling# choisir "sampling"dansla listelocal({pkg <

+ if(nchar(pkg)) library(pkg, character.only=TRUE)})
# Utilisation des donn´eesdata(swissmunicipalities)attach(swissmunicipalities)# boxplot de la s´election de

Exemple 2.10 On utilise une base de donn´ees belges fournie par l'Institut National (b elge) de Statistique contenant des infirmations sur la population et les revenus des p ersonnes physiques dans les communes. On s'int´eresse `a la variable "revenu moyen en euros par habitant en 2004" pour chaque commune (variable rageincome) et l'on aimerait comparer les 9 provinces b elges : Anvers, Brabant, Flandre occidentale, Flandre orientale,Hainaut, Li`ege, Limboug, Luxembourg, Namur. La Figure 2.6 contientles boˆıtes`amoustachesdes noms des provincesb=list("Anv."=averageincome[Province==1],"Brab."=averageincome[Province==2],"Fl.

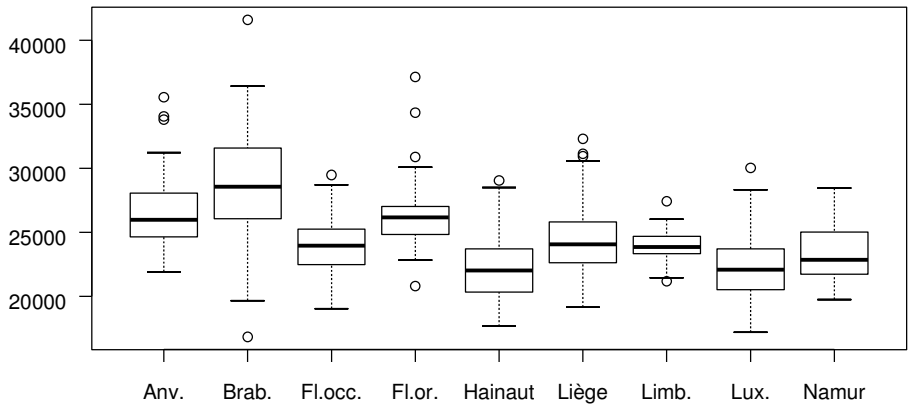


Fig. 2.6 – Boˆıtes `a moustaches du “revenu moyen des habitants” des communes selon les provinces belg

"Namur"=averageincome[Province==9])boxplot(b)**Exercices****Exercice2.1** On p`ese les 50 ´el`eves d’

esumés dans le tableau
suivant : 43 43 43 47 4848 48 48 49 4949 50 50 51 5152 53 53 53 5454 56 56 56 5759 59 59 62 6263 63 6
classes.**Exercice 2.3** 33

1. Montrez que

$$s_x^2 = \frac{1}{2n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} (x_i - x_j)^2.$$

2. Montrez que

$$s_x \leq E_t \sqrt{\frac{n-1}{2n}}.$$

3. Montrez que, si $x_i > 0$,

$$\frac{1}{n} \sum_{i=1}^{n} |x_i - \bar{x}| \leq 2\bar{x}.$$

Chapitre3

Statistique descriptive bivari´ee

3.1 S´eriestatistique bivari´ee

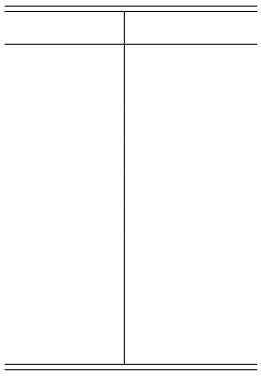
On s'int´eresse `a deux variables x et y. Ces deux variables sont mesur´ees sur un ´echantillon de n unit´es d'observation. Pourchaque unit´e,on obtient donc deux mesures et la s´eriestatistique est alors une suite de n couplesdes valeurs prises par les deux variables sur chaque individu :(x_1, y_1), ..., (x_i, y_i), ..., (x_n, y_n).

Chacune des deux variables peut ˆetre, soit quantitative, soit qualitative. On examine deux cas.

– Les deux variablessont quantitatives.– Les deux variablessont qualitatives.

3.2 Deux variables

r´eels) peut toujoursˆetre repr´esent´e comme un point dans un plan$(x_1,y_1),....,(x_i,y_i),....,(x_n,y_n)$.**Exemple 3.**



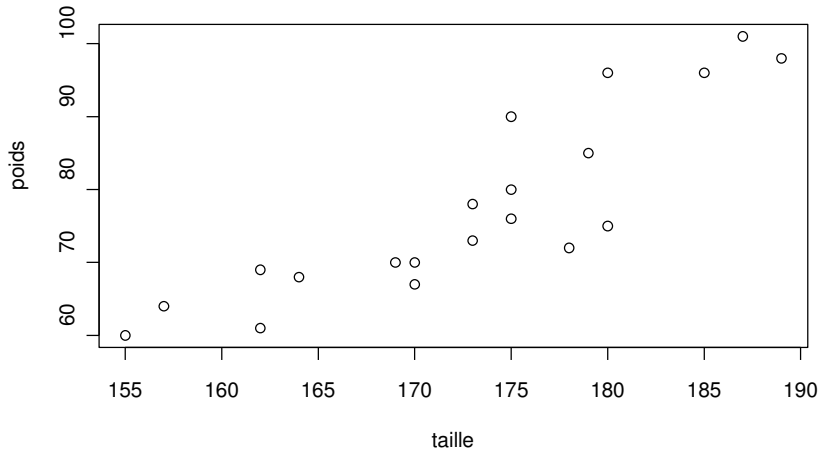


Fig. 3.1– Lenuagedepoints

3.2.2 Analyse des variables Les variables x et y peuventˆetre analys´ees s´epar´ement. On peut c

etres dont les

moyennes et les variances : $\bar{x} = \frac{1}{n}\sum_{i=1} x_i$, $s^2_x = \frac{1}{n}\sum_{i=1}(x_i - \bar{x})^2$,

$\bar{y} = \frac{1}{n}\sum_{i=1} y_i$, $s^2_y = \frac{1}{n}\sum_{i=1}(y_i - \bar{y})^2$.Ces param`etres sont appel´es *param`etres marginaux* : *variances ma*

ecarts-types

marginaux, quantiles marginaux, etc...**3.2.3 Covariance**La *covariance* est d´efinie $s_{xy} = \frac{1}{n}\sum_{i=1}(x_i - \bar{x}$

D´emonstration

$$\begin{aligned} s_{xy} &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \\ &= \frac{1}{n}\sum_{i=1}^{n}(x_i y_i - y_i \bar{x} - \bar{y} x_i + \bar{y}\bar{x}) \\ &= \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} y_i \bar{x} - \frac{1}{n}\sum_{i=1}^{n} \bar{y} x_i + \frac{1}{n}\sum_{i=1}^{n} \bar{y}\bar{x} \\ &= 1\frac{}{n}\sum_{i=1}^{n} x_i y_i - \bar{y}\bar{x} - \bar{y}\bar{x} + \bar{y}\bar{x} \\ &= 1n\,{}_{ni=1}\, x_i y_i - \bar{y}\bar{x} \end{aligned}$$

✷

3.2.4 Corr´elation Le *coefficient de corr´elation* est la covariance divis´ee par les deux´ ecart-types marginaux:

$r_{xy} = s_{xy}s_x\, s_y$ .Le *coefficient de d´etermination* est le carr´e du co efficient de corr´elation:

$r^2_{xy} = s^2_{xy}s^2_x\, s^2_y$.**Remarque 3.2**– Le coefficient de corr´elation mesure la d´ependance lin´eaire entre deux

cependant avoir une d´ependance non-lin´eaire avec un coefficient de corr´elation nul.**3.2.5 Droite de**
commet (voir Figure 3.3) en utilisant la droite de r´egression pour pr´edire y_i `a partir de x_i. Les r´esidus peu

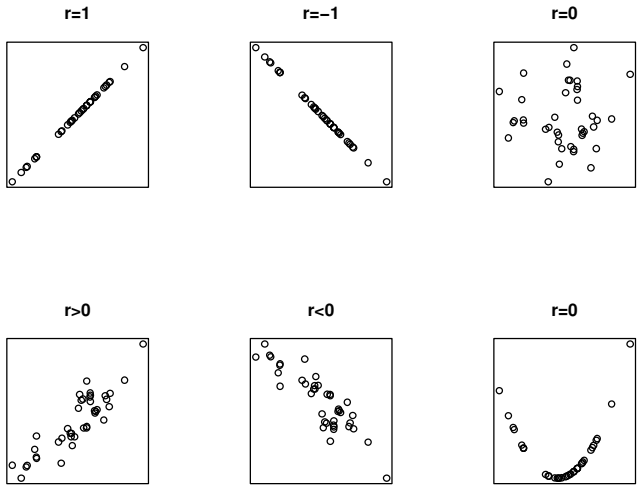


Fig. 3.2 – Exemples de nuages de points et co efficients de corr´elation

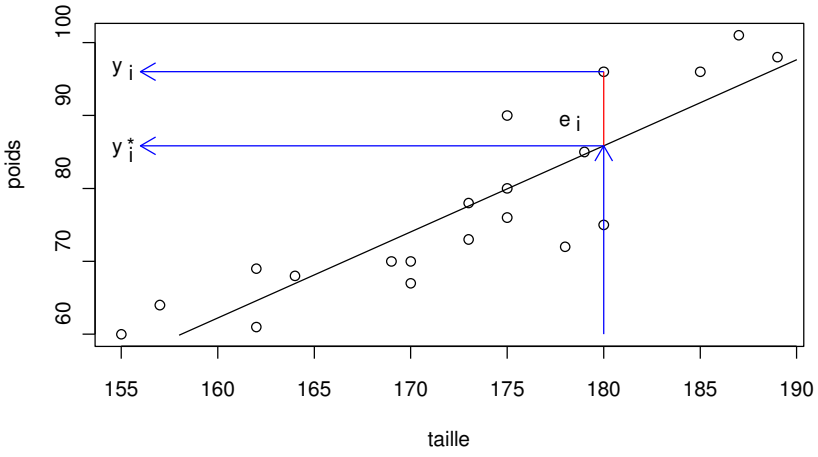


Fig. 3.3 – Le nuage de points, le r´esidu38

En langageR

```
# Graphique avec le r´esidus
plot(taille,poids)
segments(158,a+b*158,190,a+b*190)
segments(180,a+b*180,180,96,col="red")
#text(178,90,expression(e))

text(178.7,89.5,"i")#arrows(180,a+b*180,156,a+b*180,col="blue",length=0.14)


arrows(180,60,180,a+b*180,col="blue",length=0.14)
arrows(180,96,156,96,col="blue",length=0.14)
#text(154.8,86,expression(y))text(155.5,85.5,"i")#text(154.8,97,expression(y))text(155.5,97.8,"*")text(155
```

es qui consiste `a chercher la droite qui minimise la somme des carr´es des r´esidus :$M(a, b) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - a - bx_i)^2$.**Th**

es sont donn´es par:

$b = s_{xy}s_{2x}$ *et* $a = \bar{y} - b\bar{x}$.**D´emonstration** Le minimum M (a,b) en a,b s'obtient en annulant les d´eriv´ees

et b. $\frac{\partial M(a,b)}{\partial a} = -\sum_{i=1}^{n} 2\ (y_i - a - bx_i) = 0 \frac{\partial M(a,b)}{\partial b} = -\sum_{i=1}^{n} 2\ (y_i - a - bx_i)\ x_i = 0$On obtient un sys

La premièreéquation montre que la droite passe par le point moyen $(\bar{x}, \bar{y})$. On obtient

$$a = \bar{y} - b\bar{x}.$$

En remplaçant a par $\bar{y} - b\bar{x}$ dans la seconde équation, on a

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} x_i y_i - (\bar{y} - b\bar{x})\,\bar{x} - b\,\frac{1}{n}\sum_{i=1}^{n} x_i^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{y}\bar{x} - b\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2\right) \\
&= s_{xy} - b s_x^2 \\
&= 0,
\end{aligned}$$

ce qui donne

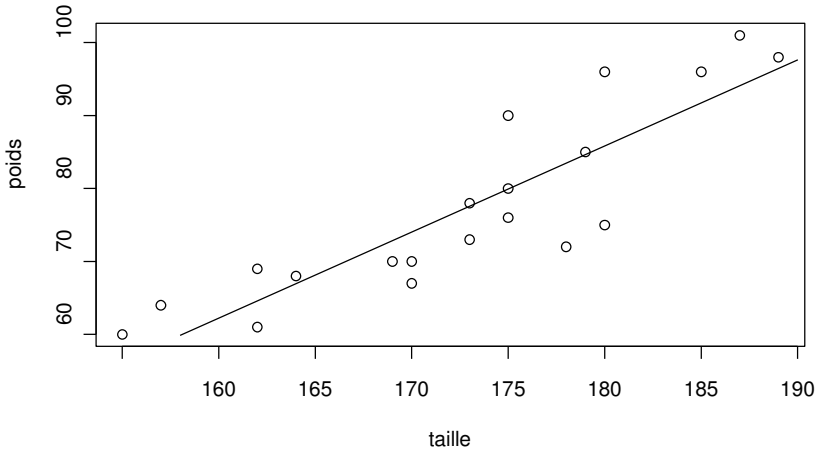
$$s_{xy} - b s_x^2 = 0.$$

Donc $b = \dfrac{s_{xy}}{s_x^2}$.

On a donc identifié les deux paramètres : $b = s_{xy}/s_x^2$ (la pente) et $a = \bar{y} - b\bar{x} = \bar{y} - (s_{xy}/s_x^2)\bar{x}$ (la constante).

On devrait en outre vérifier qu'il s'agit bien d'un minimum en montrant que les dérivées secondes sont positives.✷

La droite de régression est donc $y = a + bx = \bar{y} - (s_{xy}/s_x^2)\bar{x} + (s_{xy}/s_x^2)x$, ce qui peut s'écrire aussi $y - \bar{y} = (s_{xy}/s_x^2)(x - \bar{x})$.



Remarque 3.3 La droite de régression de y en x n'est pas la même que la droite de régression de x en y.

3.2.6 R´esidus et valeurs ajust´ees

Les *valeurs ajust´ees* sont obtenuesau moyende la droite de r´egression:

$$y_i^* = a + bx_i.$$

Les valeurs ajust´ees sont les ‘pr´edictions’ des y_i r´ealis´esau moyen de la variable x et de la droite de r´egression de y en x.**Remarque 3.4** La moyenne des valeursajust´ees est´egale `a la moyennedes valeursobserv´ees $\bar{y}$. En effet,

$$\frac{1}{n}\sum_{i=1}^{n} y_i^* = \frac{1}{n}\sum_{i=1}^{n}(a + bx_i) = a + b\frac{1}{n}\sum_{i=1}^{n} x_i = a + b\bar{x}.$$

Or, $\bar{y} = a + b\bar{x}$, car le point $(\bar{x}, \bar{y})$ appartient `a la droite de r´egression.

Les r´esidussont les diff´erencesentre les valeurs observ´ees et les valeurs ajust´ees de la variable d´ependante.

$$e_i = y_i - y_i^*.$$

Les r´esidusrepr´esententla partie inexpliqu´eedes y_i par la droite de r´egression.

Remarque 3.5– La moyenne des r´esidus est nulle. En effet$\frac{1}{n}\sum_{i=1} e_i = \frac{1}{n}\sum_{i=1}(y_i - y_{*i}) = \bar{y} - \bar{y} = 0$.

– De plus, $\sum_{i=1} x_i e_i = 0$.La d´emonstration est un peu plus difficile.

3.2.7 Sommes de carr´es et var

D´efinition 3.4 *On appel le somme des carr´es des r´esidus (ou r´esiduel le) la quantit´*

$$\mathrm{SC}_{RES} = \sum_{i=1}^{n} e_i^2.$$

D´efinition 3.5 *La variance r´esiduel le est la variance des r´esidus.*

$$s_e^2 = \frac{\mathrm{SC}_{RES}}{n} = \frac{1}{n}\sum_{i=1}^{n} e_i^2.$$

Note : Il n'est pas n´ecessaire de centrer les r´esidus sur leurs moyennes pour calculer la variance, car la moyenne des r´esidus est nulle. **Th´eor`eme 3.3** SC_{TOT}

$$= \mathrm{SC}_{REGR} + \mathrm{SC}_{RES}.$$

D´emonstration $\mathrm{SC}_{TOT} = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} (y_i - y_{*i} + y_{*i} - \bar{y})^2 = \sum_{i=1}^{n} (y_i - y_{*i})^2 + \sum_{i=1}^{n} (y_{*i} - \bar{y})^2 + 2n$

$$\sum_{i=1} (y_i - y_i^*)(y_i^* - \bar{y})$$

$= \mathrm{SC}_{RES} + \mathrm{SC}_{REGR} + 2\sum_{i=1}^{n}(y_i - y_{*i})(y_{*i} - \bar{y})$.Le troisi`eme terme est nul. En effet, $\sum_{i=1}^{n}(y_i - y_{*i})(y_{*i} - \bar{y}) = \sum_{i=}$

D´emonstration

$$\begin{aligned} s^2_{y^*} &= \frac{1}{n}\sum_{i=1}^{n}(y_i^* - \bar{y})^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}\bar{y} + \frac{s_{xy}}{s_x^2}(x_i - \bar{x}) - \bar{y}^2 \\ &= \frac{s_{xy}^2}{s_x^4}\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \\ &= \frac{s_{xy}^2}{s_x^2} \\ &= s^2_y \frac{s_{xy}^2}{s_x^2 s_y^2} \\ &= s^2_y r^2. \end{aligned}$$

✷

La *variance r´esiduel le* est la variance des r´esidus.$s^2_e = \frac{1}{n}\sum_{i=1}^{n} e_i^2$

**Th´eor`eme 3.5** *La variance r´esiduelle peut ´egalement s'´ecrire*$s^2_e = s^2_y(1 - r^2)$,*o`u* r^2 *est le coefficient d*

variables qualitatives3.3.1 Donn´ees observ´eesSi les deux variables *x* et *y* sont qua

chacune des deux variables prend comme valeurs des modalités qualitatives.

Lesvaleursdistinctesde x et y sontnotées respectivement

$$x_1, \ldots, x_j, \ldots, x_J$$

et y

$$_1, \ldots, y_k, \ldots, y_K .$$

3.3.2 Tableaude contingence

Les donn´ees observ´ees peuvent être regroupées sous la forme d'un *tableau de contingence*

	y_1	$\cdots$	y_k	$\cdots$	y_K	total
x_1	n_{11}	$\cdots$	n_{1k}	$\cdots$	n_{1K}	$n_{1.}$
$\cdots$	$\cdots$		$\vdots$		$\vdots$	
x_j	n_{j1}	$\cdots$	n_{jk}	$\cdots$	n_{jK}	$n_{j.}$
$\cdots$	$\cdots$		$\vdots$		$\vdots$	
x_J	n_{J1}	$\cdots$	n_{Jk}	$\cdots$	n_{JK}	$n_{J.}$
total	$n_{.1}$	$\cdots$	$n_{.k}$		$n_{.K}$	n

Les $n_{j.}$ et $n_{.k}$ sont appel´es les effectifs marginaux. Dans ce tableau,

– $n_{j.}$ repr´esente le nombre de fois que la modalit´e x_j apparaît,

– $n_{.k}$ repr´esente le nombre de fois que la modalit´e y_k apparaît,

– n_{jk} repr´esente le nombre de fois que

On a les relations $\sum_{j=1}^{J} n_{jk} = n_{.k}$, pour tout $k = 1,\ldots,K$, $\sum_{k=1}^{K} n_{jk} = n_{j.}$, pour tout $j = 1,\ldots,J$, et $\sum_{j=1}^{J} n_{j.} = \sum_{k=1}^{K} n_{.k} = \sum_{j=1}^{J}\sum_{k=1}^{K} n$

Le Tableau 3.1 reprend le tableau de contingence. Tab. 3.1 – Tableau des effectifs n_{jk} Bleu Vert Marron Total H

$$f_{.k} = \frac{n_{.k}}{n},\ k = 1, ..., K.$$

Le tableaudes fréquences est

	y_1	$\cdots$	y_k	$\cdots$	y_K	total
x_1	f_{11}	$\cdots$	f_{1k}	$\cdots$	f_{1K}	$f_{1.}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$	
x_j	f_{j1}	$\cdots$	f_{jk}	$\cdots$	f_{jK}	$f_{j.}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$	
x_J	f_{J1}	$\cdots$	f_{Jk}	$\cdots$	f_{JK}	$f_{J.}$
total	$f_{.1}$	$\cdots$	$f_{.k}$		$f_{.K}$	1

Exemple 3.3 LeTableau3.2reprend letableaudesfréquences.

Tab. 3.2– Tableau des fréquences

	Bleu	Vert	Marron	Total
Homme	0.05	0.25	0.10	0.40
Femme	0.10	0.30	0.20	0.60
Total	0.15	0.55	0.30	1.00

3.3.4 Profils lignes et profils colonnes Untableau de contingences'interpr`etetoujours en comp

equences

en colonnes (appel´es aussi *profils lignes* et *profils colonnes*).Les profils lignes sont d´efinis par$f_{(j)k} = n_{jk}n_{j.} =$

—— ——

—— ——

Tab. 3.4 – Tableaudes profils colonnes

	Bleu	Vert	Marron	Total
Homme	0.33	0.45	0.33	0.40
Femme	0.67	0.55	0.67	0.60
Total	1.00	1.00	1.00	1.00

3.3.5 Effectifs th´eoriqueset khi-carr´

On cherche souventune interactionentre des lignes et des colonnes, c'est-`a-dire un lien entreles variables. Pour mettre en ´evidence ce lien, on construit un tableau d'effectifs th´eoriques qui repr´esente la situation o`u les variables ne sont pas li´ees (ind´ep endance). Ces *effectifs th´eoriques* sont construits de la mani`ere suivante:

$$n^*_{jk} = \frac{n_{j.}\, n_{.k}}{n}.$$

Les effectifs observ´es $n_{j\,k}$ ont les m^emes marges que les effectifs th´eoriques n^*_{jk}.

Enfin, les *´ecarts `a l'ind´ependance* sont d´efinis par $e_{jk} = n_{j\,k} - n*$

$_{j\,k}$.

– La d´ep endance du tableau se mesure au moyen de l'indicateur d'effectif par

$$\chi^2_{obs} = \sum_{k=1}^{K}\sum_{j=1}^{J} \frac{(n_{jk} - n*_{jk})^2}{n*_{jk}} = \sum_{k=1}^{K}\sum_{j=1}^{J} \frac{e^2_{jk}}{n^*_{jk}}. \qquad (3.1)$$

– Le khi-carr´e p eut ^etre normalis´e p our ne plus d´ep endre du nombre d'observations. On définit le khi-deux par : $\varphi^2 = \chi^2_{obs}/n$ .Le φ^2 ne d´epend plus du nombre d'observations. Il est possible de montrer que $\varphi^2 \leq$ tableau ait le m^eme nombre de lignes quede colonnes).**Exemple 3.5** Le Tableau 3.5 reprend le tableau de

Tab. 3.6 – Tableau des écarts `a l'ind´endance e_{jk}

	Bleu	Vert	Marron	Total
Homme	-2	6	-4	0
Femme	2	-6	4	0
Total	0	0	0	0

Tab. 3.7 – Tableaudes e_{jk}^2 / n_{jk}^*

	Bleu	Vert	Marron	Total
Homme	0.33	0.82	0.67	1.82
Femme	0.22	0.55	0.44	1.21
Total	0.56	1.36	1.11	3.03

– Le khi-carr´e observ´e vaut $\chi^2_{obs} = 3.03$.– Le phi-deux vaut $\varphi^2 = 0.01515$.– Comme le tableau a deux li

egal

au φ^2 .– On a $V = 0.01515$. La d´ep endance entre les deux variables est tr`

es faible.

Enlangage Ryeux= c(rep("bleu",times=10),rep("vert",times=50),rep("marron",times=20),

rep("bleu",times=20),rep("vert",times=60),rep("marron",times=40))sexe= c(rep("homme",times=80),rep("
Tableaudecontingence: effectifs n_{jk}Niveaud'instruction Statutprofessionneldufilsdufilsparrapport parrapporta

Tab. 3.9 – Tableau des fréquences f_{jk}

$X\backslash Y$	Plus élevé	Egal	inférieur	total
plus élevé	0.322	0.231	0.147	0. 700
égal	0.055	0.079	0.058	0. 192
inférieur	0.017	0.038	0.053	0. 108
total	0.394	0.349	0.257	1. 000

Tab. 3.10– Tableau des profils lignes

$X\backslash Y$ Plus´	elevé	Egal	inférieur	total
plus ´elev´e 0.460		0. 330	0. 210	1
´egal 0. 288		0. 413	0. 300	1
inf´erieur 0. 156		0. 356	0. 489	1
total 0.	394	0. 349	0. 257	1

Tab. 3.11 – Tableaudes profils colonnes

$X\backslash Y$ Plus ´elev´e	Egal	inf´erieur	total
plus ´elev´e 0.817	0.662	0.570	0. 700
´egal 0.140	0 .228	0.224	0. 192
inf´erieur 0.043	0 .110	0 .206	0. 108
total 1	111		

Tab. 3.12 – Tableau des effectifs th´eoriques $n_{*jk}$$X\backslash Y$ Plus ´elev´e Egal inf´erieur totalplus ´elev´e 114.72 1.24 12.52 χ^2_{obs} = 25.81On a donc χ^2_{obs} = 25.81$\varphi^2 = \chi^2_{obs}n$ = 25.81416 = 0.062$V = \varphi^2\min(J-1, K-1)$ =

Exercices**Exercice 3.1** Laconsommationdecr`

emes glaces par individus a été mesurée pendant 30 périodes. L'objectif est d´etermin´e si la consommation moyenne d´epend de la temp´erature. Les donn´ees sont dans le tableau 3.15. Onsaiten outreque Tab. 3.15–Consommationde cr`

emes glaces

consommation y	temp´erature x	consommation y	temp´erature x	consommation y	temp´erature x
386	41	286	28	319	44
374	56	298	26	307	40
393	63	329	32	284	32
425	68	318	40	326	27
406	69	381	55	309	28
344	65	381	63	359	33
327	61	470	72	376	41
288	47	443	72	416	52
269	32	386	67	437	64
256	24	342	60	548	71

$\sum_{i=1}^{n} y_i = 10783$, $\sum_{i=1}^{n} x_i = 1473$, $\sum_{i=1}^{n} y_i^2 = 4001293$, $\sum_{i=1}^{n} x_i^2 = 80145$, $\sum_{i=1}^{n} x_i y_i = 553747$,1. Donnez les moyenne

2. Donnez la droite de r´egression, avec comme variable d´ep endante la consommation de glaces et co
variableexplicativelatemp´erature.3. Donnezlavaleurajust´eeetler´esidupourlapremi`ereobservationdutable
´etationdececoefficient.3. D´eterminezladroiteder´egressionY enfonctiondeX.4. ´Etablissez,surbasedumod`

Exercice 3.3 Consid´erons un´echantillonde 10 fonctionnaires (ayant entre 40 et 50 ans) d'un certain minist`ere.
Soit X le nombre d'années de service et Y le nombre de jours d'absence pour raison de maladie (au cours de l'ann´ee pr´ec´edente) et d´etermin´e pour chaque p ersonne appartenant `a cet´echantillon.

x_i	2	14	16	8	13	20	24	7	5	11
y_i	3	13	17	12	10	8	20	7	2	8

1. Repr´esentez le nuage de points.
2. Calculez le co efficient de corr´elation entre X et Y.
3. D´eterminez l'´equation de la droite de r´egression de Y en fonction de X.
4. D´eterminez la qualit´e de cet ajustement.
5. ´Etablissez, sur base de ce mo d`ele, le nombre de jours d'absence p our un fonctionnaire ayant 22 ans de service. 50

Chapitre4

Th´eorie des indices, mesures d'in´egalit´e

4.1 Nombres indices

4.2 D´efinition

Un indice est la

e´erence. Prenons l'exemple du tableau
4.1 contenant le prix (fictif) d'un bien de consommation de 2000 `a 2006. Le temps varie de 0, 1, 2, .. ., 6 e
est consid´er´e comme le temps de r´ef´erence par rapp ort auquel l'indice est calcul´

Tab. 4.1– Tableau du prix d'un bien de consommation de 2000 `a 2006

ann´ee t prix p_t2000 0 2.002001 1 2.302002 2 2.402003 3 2.802004 4 3.002005 5 3.502006 6 4.00L'indice
142.864 66.67 76.67 80.00 93.33 100.00 116.67 133.335 57.14 65.71 68.57 80.00 85.71 100.00 114.296 5

4.2.1 Propriétés des indices

Considérons un indice quelconque $I(t/0)$. On dit que cet indice possède les propriétés de
– *réversibilité* si $I(t/0) = 100^2 \times \frac{1}{I(0/t)}$,
– *identité* si $I(t/t) = 100$,
– *circularité (ou transitivité)* si $I(t/u) \times I(u/v) = 100 \times I(t/v)$.
Il est facile de montrer que ces quatre propriétés sont satisfaites pour un indice simple.

4.2.2 Indices synthétiques

Quand on veut calculer un indice à partir de plusieurs prix, le problème devient sensiblement plus compliqué. Un indice synthétique est une grandeur d'un ensemble de biens par rapport à celle d'une année référence. On ne peut pas construire un indice synthétique en additionnant simplement des indices simples. Il faut, en effet, tenir compte des quantités achetées.

Pour calculer un indice de prix de n biens de consommation étiquetés de 1, 2, .. ., n, on utilise la notation suivante: – p_{ti} représente le prix du bien de consommation i au temps t,

– q_{ti} représente la quantité de biens i consommé au temps t.
Considérons par exemple le Tableau 4.3 qui contient 3 biens de consommation et pour lesquels ont connu les prix et les quantités achetées. Tab. 4.3 – Exemple : prix et quantités de trois bien pendant 3 ans

Temps	0 12			
	Prix ($p_{0\,i}$) Quantités ($q_{0\,i}$)	Prix ($p_{1\,i}$) Quantités (q_{1i})	Prix (p_{2i})	Quantités (q_{2i})
Bien 1	100 14	150 10	200	8
Bien 2	60 10	50 12	40	14
Bien 3	160 4	140 5	140	5

Il existe deux méthodes fondamentales pour calculer les indices de prix, l'indice de Paasche et l'indice de Laspeyres.

4.2.3 Indice de Laspeyres

L'indice de Laspeyres, est défini par $L(t/0) = 100 \times \sum_{i=1}^{n} q_{0i} p_{ti} \sum_{i=1}^{n} w_{0i} I_i(t) \sum_{i=1}^{n} w_{0i} = \sum_{i=1}^{n} p_{0i} q_{0i} 100 \times \sum_{p_{ti} p_{0i} i=1}^{n} p_{0i} q_{0i} = 100 \times \sum_{i=1}^{n} q_{0i} p_{ti} \sum_{i=1}^{n} p_{0i} q_{0i}$. L'indice de Laspeyres ne possède

Exemple 4.1 Sionutiliselesdonnées du tableau 4.3, les indices de Lasp eyres sont les suivants

$$L(1/0) = 100\times\frac{\sum_{i=1}^{n} q_{0i}p_{1i}}{\sum_{i=1}^{n} q_{0i}p_{0i}} = 100\times\frac{14\times 150 + 10\times 50 + 4\times 140}{14\times 100 + 10\times 60 + 4\times 160} = 119.6970,$$

$$L(2/0) = 100\times\frac{\sum_{i=1}^{n} q_{0i}p_{2i}}{\sum_{i=1}^{n} q_{0i}p_{0i}} = 100\times\frac{14\times 200 + 10\times 40 + 4\times 140}{14\times 100 + 10\times 60 + 4\times 160} = 142.4242,$$

$$L(2/1) = 100\times\frac{\sum_{i=1}^{n} q_{1i}p_{2i}}{\sum_{i=1}^{n} q_{1i}p_{1i}} = 100\times\frac{10\times 200 + 12\times 40 + 5\times 140}{10\times 150 + 12\times 50 + 5\times 140} = 113.5714.$$

4.2.4 Indice de Paasche

L'indice de Paasche, est d´efini par $P(t/0) = 100\times$

$$\frac{\sum_{i=1}^{n} q_{ti}p_{ti}}{\sum_{i=1}^{n} q_{ti}p_{0i}}.$$

On utilise, pour le calculer, les quantit´es q_{ti} du temps par rapp ort auquel on veut calculer l'indice.

L'indice de Paasche peut aussi ˆetre pr´esent´e comme une moyenne harmonique des indices
simples. Soient l'indice simple du bien i : $I_i(t/0) = 100\times p_{ti}p_{0i}$,et le poids w_{ti} correspondant `a la recette tot

les recettes au temps t :$P(t/0) = \sum_{i=1}^{n} w_{ti}\sum_{i=1}^{n} w_{ti}/I_i(t/0) = \sum_{i=1}^{n} p_{ti}q_{ti}\sum_{i=1}^{n} p_{ti}q_{ti}\,p_{0i}100\times p_{ti} = 100\times\sum_{i=1}^{n} q_{ti}p_{ti}\sum_{i=1}^{n} q_{ti}p_{0i}$ .I

difficile `a calculer que l'indice de Laspeyres, car on doit connaˆıtre les quantit´es pour chaque valeur de t.

Fisher L'indice de Laspeyres est en g´en´eral plus grand que l'indice de Paasche, ce qui peut s'expliquer p

Fisher a propos´e d'utiliser un compromis entre l'indice de Paasche et de Lasp eyres en calculant simplement la moyenne g´eom´etrique de ces deux indices

$$F(t/0) = \sqrt{L(t/0) \times P(t/0)}.$$

L'avantage de l'indice de Fisher est qu'il jouit de la propri´et´e de r´eversibilit´e.

Exemple 4.3 Si on utilise toujours les donn´ees du tableau 4.3, les indices de Fisher sont les suivants:

$$F(1/0) = \sqrt{L(1/0) \times P(1/0)} = 115.3242,$$

$$F(2/0) = \sqrt{L(2/0) \times P(2/0)} = 129.2052,$$

$$F(2/1) = \sqrt{L(2/1) \times P(2/1)} = 111.7715.$$

4.2.6 L'indice de Sidgwick

L'indice de Sidgwick est la moyenne arithm´etique des indices de Paasche et de Laspeyres:

$$S(t/0) = \frac{L(t/0) + P(t/0)}{2}.$$

4.2.7 Indices cha^ınes

Le d´efaut principal des indices de Laspeyres, de Paasche, de Fisher et de Sidgwick est qu'ils ne poss`edent pas la propri´et´e de circularit´e. Un indice qui poss`ede cette propri´et´e est appel´e indice cha^ıne. Pour construire un indice cha^ıne, avec l'indice de Lasp eyres, on peut faire un produit d'indice de Lasp eyre annuels.

$$CL(t/0) = 100 \times \frac{L(t/t-1)}{100} \times \frac{L(t-1/t-2)}{100} \times \cdots \times \frac{L(2/1)}{100} \times \frac{L(1/0)}{100}.$$

Pour calculer un tel indice, on doit ´evidemment connaître les quantit´es pour chaque valeur de t. L'indice suisse des prix `a la consommation est un indice cha^ıne de Laspeyres. **Exemple 4.4** En utilisant encore les

suivants : $CL(1/0) = L(1/0) = 119.6970$, $CL(2/1) = L(2/1) = 113.5714$, $CL(2/0) = L(2/1) \times \frac{L(1/0)}{100} = 135.9416$. **4**

4.3.2 Courbe de Lorenz

Plusieurs indices d'inégalité sont liés à la courbe de Lorenz. On note

$$x_1, \ldots, x_i, \ldots, x_n$$

les revenus des n individus de la population étudiée. On note également

$$x_{(1)}, \ldots, x_{(i)}, \ldots, x_{(n)},$$

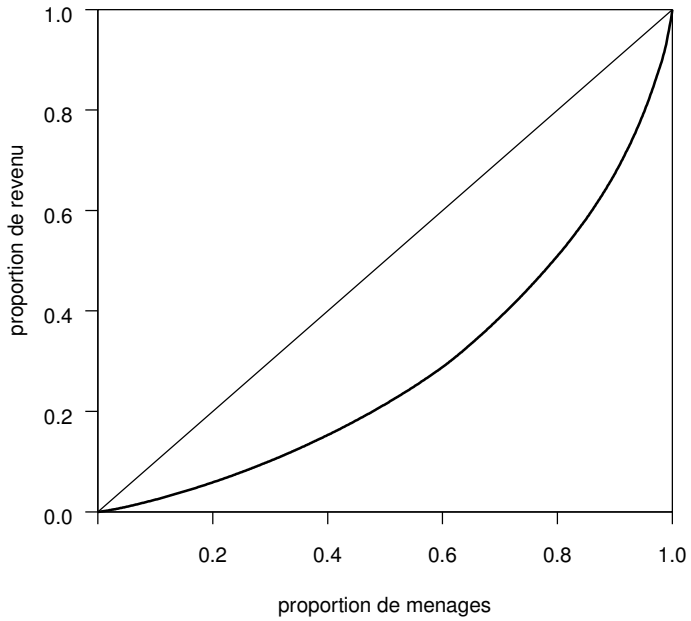
la statistique d'ordre, c'est-à-dire la série des revenus triés par ordre croissant.

Notons maintenant q_i la proportion de revenus par rapport au revenu total que reçoivent les i individus ayant les plus bas revenus, ce qui s'écrit

$$q_i = \frac{\sum_{j=1}^{i} x_{(j)}}{\sum_{j=1}^{n} x_{(j)}} \text{ avec } q_0 = 0 \text{ et } q_n = 1.$$

La courbe de Lorenz est la représentation graphique de la fonction qui à la part des individus les moins riches associe la part y du revenu total qu'ils perçoivent. Plus précisément, la courbe de Lorenz relie les points $(i/n, q_i)$ pour $i = 1, .., n$. En abscisse, on a donc une proportion d'individus classés par ordre de revenu, et en ordonnée la proportion du revenu total reçu par ces individus.

Exemple 4.5 On utilise une enquête ménage sur le revenu des ménages aux Philippines appelée Cette enquête de 1997 sur le revenu des ménages a été réalisée par l'Office philippin de la Statistique. La courbe de Lorenz est présentée en Figure 4.1. Fig. 4.1 – Courbe de Lorenz



Remarque 4.1 Sur le graphique, on indique toujours la diagonale. La courbe de Lorenz est égale à la diagonale si tou

##CourbedeLorenzetindicesd'in´egalit´e

##Etape1:oninstallelapackageineq

utils:::menuInstallPkgs()
#choisir'ineq'danslaliste
##Etape2:onchargelepackageineq

local({pkg<-select.list(sort(.packages(all.available=TRUE)))
+if(nchar(pkg))library(pkg,character.only=TRUE)})
#choisir'ineq'danslaliste
##Utilisationdelabasededonn´eesIlocos

#Enqu^etesurlerevenudel'OfficedeStatistiquePhilippin
data(Ilocos)attach(Ilocos)#plot(Lc(income),xlab="proportiondemenages",

ylab="proportionderevenu",main="")**4.3.3IndicedeGini**L'indicedeGini,not´eGest´egal`adeuxfoislasurf

Ilestpossibledemontrerque:$G=\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i-x_j|2^{-}x.$Enutilisantlastatistiqued'ordre$x_{(1)},...,x_{(i)},...,x_{(n)}$,l'indiced

ecrire

$G=\frac{1}{n}-\frac{2}{n}\sum_{i=1}^{n}ix_{(i)}n^{-}x-(n+1).$L'indicedeGiniestcomprisentre0et1.S'ilestprochede0,touslerevenussont´egaux.S

1,lesrevenussonttr`esin´egaux.**4.3.4IndicedeHoover**L'indiced'´equir´epartitiondeHoover(ou*RobinHo*
´e`alacourbedeLorenz,carilestpossibledemontrerqu'ilcorrespond`alaplusgrandedistanceverticaleentrelacour

Lequintileshareratioestdéfinitpar

$$QSR = \frac{S_{80}}{S_{20}}.$$

Ledecileshareratioestdéfinitpar

$$DSR = \frac{S_{90}}{S_{10}}.$$

Cesquantit´essonttoujoursplusgrandesque1etaugmententavecl'in´egalit´e. Cesdeuxrapportssont facilementinterpr´etables,parexemplesi$QSR = 5$, celasignifiequelerevenumoyendes20%lesplus richesest5foisplusgrandquelerevenumoyende20%lespluspauvres.

4.3.6 Indicedepauvret´e

Unindicesimpledepauvret´econsiste`acalculerlepourcentagedelapopulationg

moiti´edelam´ediane. **4.3.7 Indicesselonlespays** Letableau4.4reprendpourtouslespaysl'indicedeGini

etlerapportdes20%lesplusrichessurles 20%lespluspauvres.(r´ef´erence:UnitedNations2005DevelopmentProgrammeReport,page270).

Exercices

Exercice4.1 ´Etudiezlespropri´et´es(circularit´e,r´eversibilit´e,identit´

eettransitivité)detouslesindicesde prixpr´esent´es.57

Tab.4.4 – Mesures de l'inégalité par pays

Rang	Pays	Indice de Gini	DSR	QSR	Année de l'enquête
1	Denmark	24.7	8.1	4.3	1997
2	Japan	24.9	4.5	3.4	1993
3	Sweden	25	6.2	4	2000
4	Belgium	25	7.8	4.5	1996
5	CzechRepublic	25.4	5.2	3.5	1996
6	Norway	25.8	6.1	3.9	2000
7	Slovakia	25.8	6.7	4	1996
8	BosniaandHerzegovina	26.2	5.4	3.8	2001
9	Uzbekistan	26.8	6.1	4	2000
10	Finland	26.9	5.6	3.8	2000
11	Hungary	26.9	5.5	3.8	2002
12	RepublicofMacedonia	28.2	6.8	4.4	1998
13	Albania	28.2	5.9	4.1	2002
14	Germany	28.3	6.9	4.3	2000
15	Slovenia	28.4	5.9	3.9	1998
16	Rwanda	28.9	5.8	4	1983
17	Croatia	29	7.3	4.8	2001
18	Ukraine	29	6.4	4.3	1999
19	Austria	30	7.6	4.7	1997
20	Ethiopia	30	6.6	4.3	1999
21	Romania	30.3	8.1	5.2	2002
22	Mongolia	30.3	17.8	9.1	1998
23	Belarus	30.4	6.9	4.6	2000
24	Netherlands	30.9	9.2	5.1	1999
25	Russia	31	7.1	4.8	2002
26	SouthKorea	31.6	7.8	4.7	1998
27	Bangladesh	31.8	6.8	4.6	2000
28	Lithuania	31.9	7.9	5.1	2000
29	Bulgaria	31.9	9.9	5.8	2001
30	Kazakhstan	32.3	7.5	5.1	2003
31	Spain	32.5	9.5	4.1	990
32	India	32.5	7.3	4.9	1993
33	Tajikistan	32.6	7.8	5.2	2003
34	France	32.7	9.1	5.6	1995
35	Pakistan	33	7.6	4.8	1998
a		57.8	33.1	17.9	2000
117	Brazil	59.3	68	26.4	2001
118	Guatemala	59.9	55.1	24.4	2000
119	Swaziland	60.9	49.7	23.8	1994
120	CentralA				

Chapitre5

Sériestemporelles,filtres,moyennes mobilesetdésaisonnalisation

5.1Définitionsgénéralesetexemples

5.1.1Définitions

Définition5.1 *Unesérietemporelleestunesuited'observationsd'unequantitérépétéedansletemps.*

Onénonceengénérall'hypothèsequelesintervallesdetempssontéquidistants.Lasérietemporelleestnotée $y_1,\ldots,y_t,\ldots,y_T$.Onnoteégalement $T=\{1,2,\ldots,t,\ldots,T\}$ l'ensembledesinstantsauxquelslesobservationssontréalisées.

Unesérietemporellepeutsecomposerde:–unetendanceT_t,–unecomposantecycliqueC_t(nousn'étudieronspasc
isoleretestimerunecomposantesaisonnière,etdésaisonnaliserlasérie,–réaliserunepr´evisionpourdesvaleu
QTR:Quarter,trimestresdepuisle1ertrimestre1978jusqu'au4ètrimestre198559

–DISH:Nombredelave-vaisselles(dishwashers)exp´edi´es(milliers)
–DISP:Nombredebroyeurd'ordures(disposers)exp´edi´es(milliers)
–FRIG:Nombreder´efrig´erateursexp´edi´es(milliers)
–WASH:Nombredemachines`alaver(washingmachine)exp´edi´es(milliers)
–DUR:D´epensesenbiensdurablesUSA(milliardsdedollarsde1982)
–RES:Investissementr´esidentielprivéUSA(milliardsdedollarsde1982)

Tab.5.1–BiensmanufacturésauxUSA

QTR	DISH	DISP	FRIG	WASH	DUR	RES
1	841	798	1317	1271	252.6	172.9
2	957	837	1615	1295	272.4	179.8
3	999	821	1662	1313	270.9	180.8
4	960	858	1295	1150	273.9	178.6
5	894	837	1271	1289	268.9	174.6
6	851	838	1555	1245	262.9	172.4
7	863	832	1639	1270	270.9	170.6
8	878	818	1238	1103	263.4	165.7
9	792	868	1273		260.6	154.9
10	589	623	1031		231.9	124.1
11	656	766	1143		242.7	126.8
12	699	822	1118	1248.6		142.2
13	675	871	1196	1258.7		139.3
14	652	791	1141	1248.4		134.1
15	628	759	1411	1255.5	122.3	
16	529	734	9191	1240.4	110.4	
17	480	706	9431	1036	247.7	101.2
18	530	582	1175	1019	249.1	103.4

195576591269104725 1.8 100.1 20602837973918262.0 115.8 216588671102

´erateursvendusamanifestementunecomposantesaisonni`ereetunetendance.**EnlangageR**QTR=c(1,2,3,4,5,

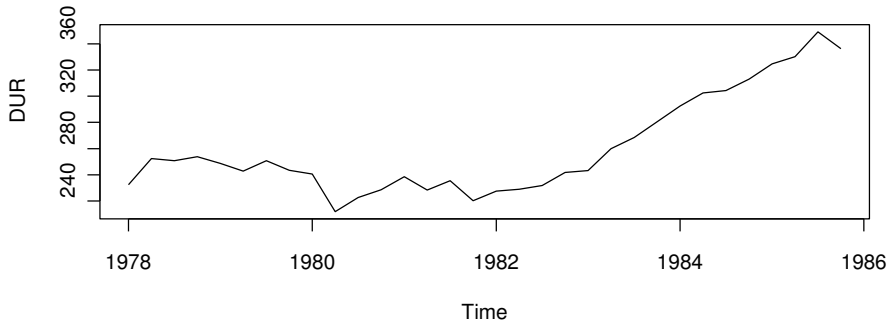


Fig.5.1–D´epensesenbiensdurablesUSA(milliardsdedollarsde1982)

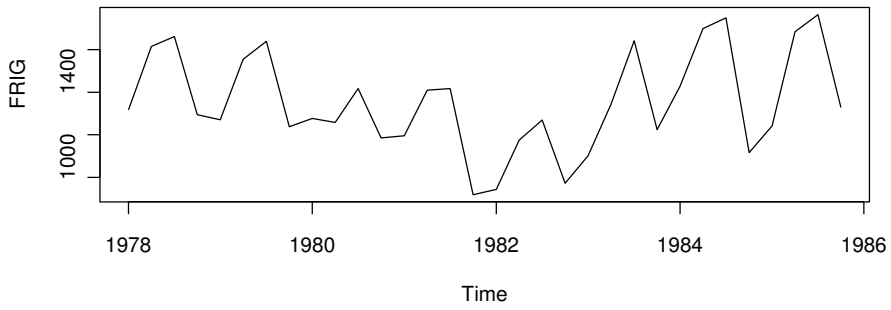


Fig.5.2–Nombreder´efrig´erateursvendusde1978`a1985

DISP=c(798,837,821,858,837,838,832,818,868,623,662,822,871,791,759,734,706,
582,659,837,867,860,918,1017,1063,955,973,1096,1086,990,1028,1003)
FRIG=c(1317,1615,1662,1295,1271,1555,1639,1238,1277,1258,1417,1185,1196,1410,1417,919,943,1175

1242,1684,1764,1328)WASH=c(1271,1295,1313,1150,1289,1245,1270,1103,1273,1031,1143,1101,1181,1
4,161.9,159.9,170.5,173.1,170.3,169.6,170.3,172.9,175,179.4)plot(QTR,DUR,type="l")plot(QTR,FRIG,type=

Tab.5.2–Indicedesprix`alaconsommation(France)

p_t	1970	1971	1972	1973	1974	1975	1976	1977	1978
janvier	97.9	102.5	108.3	115.5	127.4	145.9	159.9	174.3	190.3
f´evrier98.2		103.0	108.9	115.8	129.1	147.0	161.0	175.5	191.7
mars98.5		103.4	109.4	116.4	130.6	148.2	162.4	177.1	193.4
avril99.0		104.0	109.8	117.2	132.7	149.5	163.8	179.4	195.5
mai99.4		104.7	110.4	118.3	134.3	150.6	164.9	181.1	197.4
juin99.8105.1			111.0	119.2	135.8	151.7	165.6	182.5	198.9
juillet100.0105.6			111.9	120.2	137.5	152.8	167.2	184.1	201.5
aoˆut100.4106.0			112.5	121.0	138.6	153.8	168.4	185.1	202.5
septembre100.8106.5			113.2	122.1	140.1	155.1	170.2	186.7	203.8
octobre101.2107.1114.2				123.4	141.8	156.3	171.8	188.2	205.7
novembre101.6107.5114.9				124.5	143.1	157.3	173.2	188.9	206.8
d´ecembre101.9108.0115.5125.3					144.3	158.2	173.8	189.4	207.8

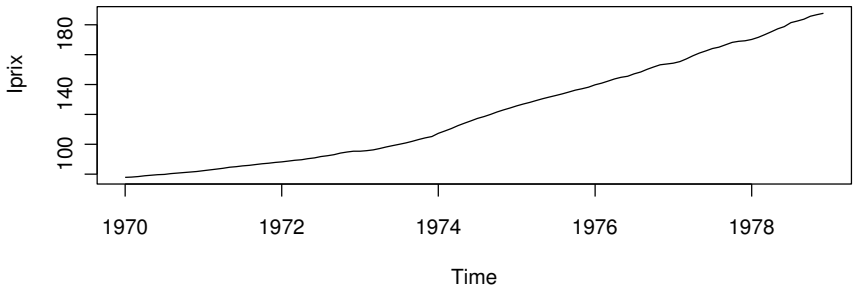


Fig.5.3–Indicedesprix`alaconsommationp_tTime

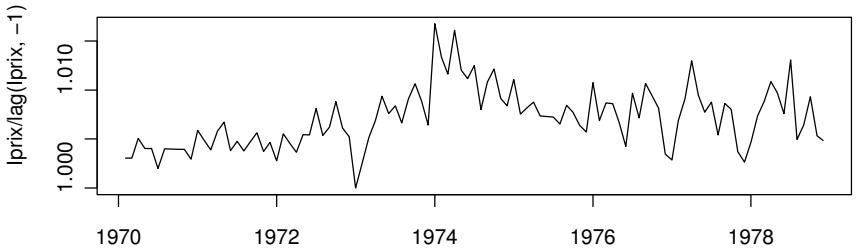


Fig.5.4–Rapportmensueldesindicesdeprixp_t/p_{t-1}Time

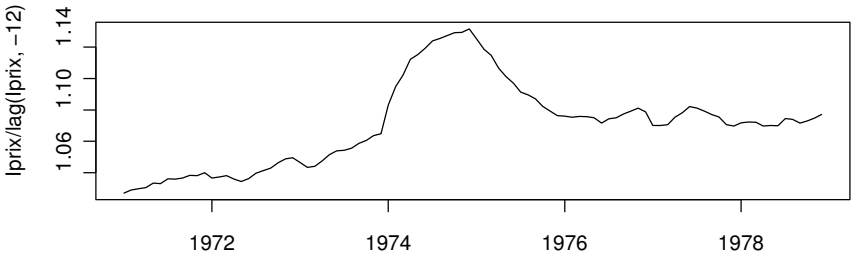


Fig.5.5–Rapportenglissementannueldesindicesdeprixp_t/p_{t-12}#lprix=c(97.9,98.2,98.5,99,99.4,99.8,100,100.4

108.3,108.9,109.4,109.8,110.4,111,111.9,112.5,113.2,114.2,114.9,115.5,
115.5,115.8,116.4,117.2,118.3,119.2,120.2,121,122.1,123.4,124.5,125.3,
127.4,129.1,130.6,132.7,134.3,135.8,137.5,138.6,140.1,141.8,143.1,144.3,
145.9,147,148.2,149.5,150.6,151.7,152.8,153.8,155.1,156.3,157.3,158.2,
159.9,161,162.4,163.8,164.9,165.6,167.2,168.4,170.2,171.8,173.2,173.8,
174.3,175.5,177.1,179.4,181.1,182.5,184.1,185.1,186.7,188.2,188.9,189.4,
190.3,191.7,193.4,195.5,197.4,198.9,201.5,202.5,203.8,205.7,206.8,207.8)
Iprix<-ts(Iprix,start=c(1970,1),frequency=12)
plot(Iprix)plot(Iprix/lag(Iprix,-1))plot(Iprix/lag(Iprix,-12))**Exemple5.4**Donn´eesdunombredevoyageurs-kilom

etresendeuxièmeclasseexprimésenmillionsde
kilom`etres.Tab.5.3–TraficdunombredevoyageursSNCF

mois/ann´eejanv.f´ev.marsavrilmaijuinjuil.	aoˆut	sept.	oct.	nov.	déc.
1963175015601820209019102410314 0	2850	2090	1850	1630	2420
1964171016001800212021002460320 0	2960	2190	1870	1770	2270
196516701640177021902020261031902860		2140	1870	1760	2360
196618101640186019902110250030302900		2160	1940	1750	2330
196718501590188022102110248028802670		2100	1920	1670	2520
19681834179218602138211524852581263920 38			1936	1784	2391
19691798185019812085212024912834272519 32			2085	1856	2553
197018541823200524182219272229122771215321 36				1910	2537
19712008183521202304226421752928273821782137				2009	2546
197220842034215225222318268429712759226721 52				1978	2723
19732081211222792661228129293089280322962210 2135					2862
19742223224824212710250530213327304426072525 2160					2876
19752481242825962923279532873598311828752754 2588					3266
1976266726682804280629763430370530532764280227073307					
19772706258627962978305334633649309528392966286333 75					
19782820285733063333314135123744317929842950289636111979331326442872326733913682					

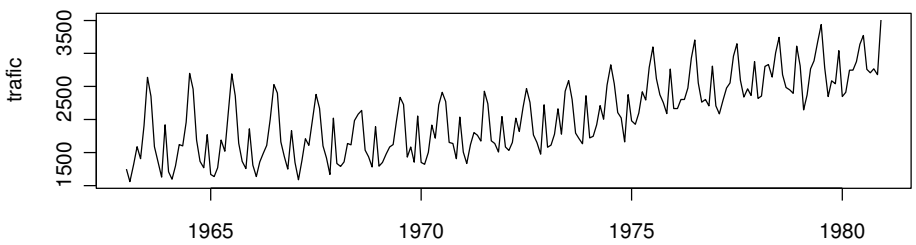


Fig.5.6–TraficdunombredevoyageursSNCF63

5.2 Description de la tendance

5.2.1 Les principaux modèles

Plusieurs types de modèles peuvent être utilisés pour décrire la tendance.

– Modèles dépendant du temps. La série dépend directement du temps. Le modèle peut être additif :

$$y_t = f(t) + E_t,$$

ou multiplicatif

$$y_t = f(t) \times E_t.$$

– Modèles explicatifs statiques : la série chronologique dépend des valeurs prises par une ou plusieurs autres séries chronologiques.

$$y_t = f(x_t) + E_t$$

Le cas linéaire est le plus facile à traiter $y_t = b_0 + b_1 x_t + E_t$.

– Modèles auto-projectifs. La série chronologique au temps t dépend de ses propres valeurs passées

$$y_t = f(y_{t-1}, y_{t-2}, y_{t-3}, \ldots, y_{t-p}) + E_t$$

– Modèles explicatifs dynamiques : la série chronologique dépend des valeurs présentes et passées d'une ou de plusieurs autres séries chronologiques, par exemple :

$$y_t = \mu + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \cdots + \theta_p y_{t-p} + \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \cdots + \varphi_q x_{t-q} + E_t.$$

5.2.2 Tendance linéaire

La tendance la plus simple est linéaire. On peut estimer les paramètres au moyen de la méthode des moindres carrés. C'est une régression simple. $T_t = a + bt$.

5.2.3 Tendance quadratique

On peut utiliser une [...] des moindres carrés. C'est une régression avec deux variables explicatives. $T_t = a + bt + ct^2$

5.2.4 Tendance po[...]

[...]éaire, on peut se ramener à un problème linéaire. En posant $z_t = 1/T_t$, on a $z_t = 1 + be^{-at}c$ 64

$$
\begin{aligned}
z_{t+1} &= \frac{1+be^{-a(t+1)}}{c} \\
&= \frac{1+be^{-at}e^{-a}}{c} \\
&= \frac{1+(1+be^{-at})e^{-a}-e^{-a}}{c} \\
&= \frac{1-e^{-a}}{c}+z_t e^{-a}.
\end{aligned}
$$

Enposant$\alpha=$

$$\frac{1-e^{-a}}{c},\ \text{et}\beta=e^{-a}.$$

onobtient$z$

$$_{t+1}=\alpha+\beta z_{t},$$

cequiestunmod`eleauto-projectif.Onpeutalorsd´eterminerlesvaleursde$\alpha$et$\beta$parunesimpler´egression lin´eaire.Ensuiteond´eduit$a$delamani`eresuivante:

$$a=-\log\beta,$$

etcomme$\alpha=1-e_{-a}c=1-\beta$

$$c,$$

ond´etermine$c$par$c=1-\beta\alpha$.Enfin,onremarqueque$z_t-1c=be_{-at}c$,onpeutd´eterminerautantdevaleursde$b$quel'

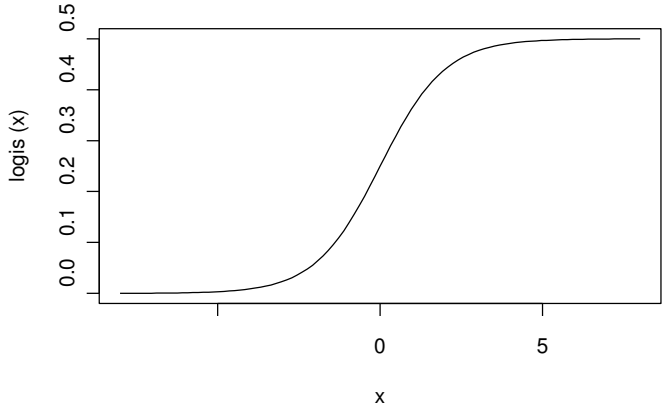


Fig.5.7–Exempledefonctionlogistiqueavec$c=0.565$

5.3 Opérateurs de décalage et de différence

5.3.1 Opérateurs de décalage

Afin de simplifier la notation, on utilise des opérateurs de décalage. On définit l'opérateur de décalage "retard" (en anglais *lag operator*) L par

$$Ly_t = y_{t-1},$$

et l'opérateur (en anglais *forward operator*) "avance" F

$$Fy_t = y_{t+1},$$

l'opérateur identité Iy

$$_t = y_t.$$

L'opérateur avance est l'inverse de l'opérateur retard

$$FL = LF = I.$$

On peut donc écrire $F^{-1} = L$ et L

$$^{-1} = F.$$

On a également – $L^2 y_t = LLy_t = y_{t-2}$, – $L^q y_t = y_{t-q}$, – $F^q y_t = y_{t+q}$, – $L^0 = F^0 = I$, – $L^{-q} y_t = F^q y_t = y_{t+q}$.

5.3.2 Opérateur diff

ée dans le graphique 5.9.

En langage R

```
## Tendance linéaire et différence
lin=10+0.3*(0:50)+rnorm(50,0,1)p
```

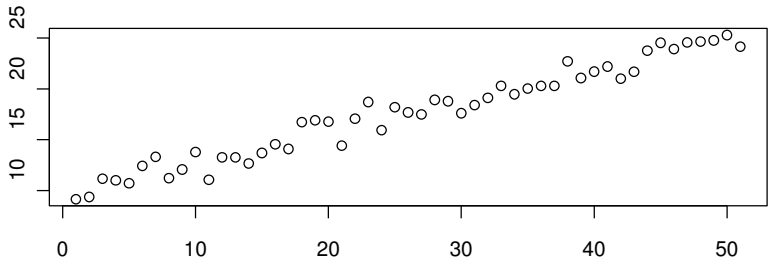


Fig.5.8–S´erieavecunetendancelinéairedépendantdutemps

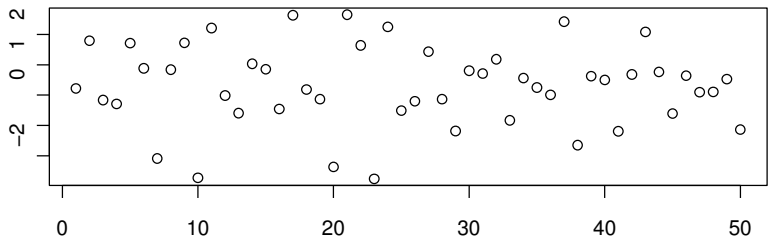


Fig.5.9–Diff´erenced'ordreundelas´erieavecunetendancelinéaire

Onpeutconstruirel'op´erateurdiff´erenced'ordredeuxen´elevant∇aucarr´

$\nabla^2 = \nabla \times \nabla = I - 2L + L^2$L'op´erateurdiff´erenced'ordredeuxpermetd'enleverunetendancequadratique.Eneffet,sila

eries'écrit

$y_t = a + b \times t + c \times t^2 + E_t$,alors$\nabla^2 y_t = (I - 2L + L^2) y_t = a + b \times t + c \times t^2 + E_t - 2a - 2b \times (t-1) - 2c \times (t-1)^2 - 2E_{t-1} + a + b \times (t-2)$
´erateurs,lacomposantesaisonni`eredispara^ıt.**EnlangageR**67

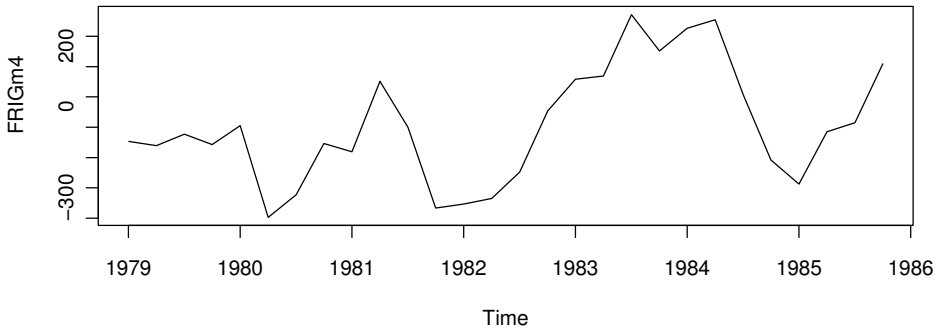


Fig.5.10–Diff´erenced'ordre4delavariableventede'r´efrig´erateurs'

##Venteder´efrig´erateursdiff´erenced'ordre4

#FRIGm4=FRIG-lag(FRIG,-4)plot(FRIGm4)**Exemple5.7**Sionappliqueunediff´erencesaisonni`ered'ordre12sur

eesdunombredevoyageurs-
kilom`etres$y_t$endeuxi`emeclasseexprim´eesenmillionsdekilom`etresdelaSNCF,ndancesaisonnier`e
dispara ̂ıt(voirFigure5.12).Onaainsilanouvellevariable$z_t=\nabla_{12}y_t=(I-L_{12})y_t=y_t-y_{t-12}$.Uneautremani`eredefaire

erence,
cequirevient`aprendrelelogarithmedurapportdelavariable(voirFigure5.13).Ond´ efinitainsi une
nouvellevariablev_t:$v_t=\nabla_{12}\log y_t=(I-L_{12})\log y_t=\log y_t-\log y_{t-12}=\log y_t y_{t-12}$.Time

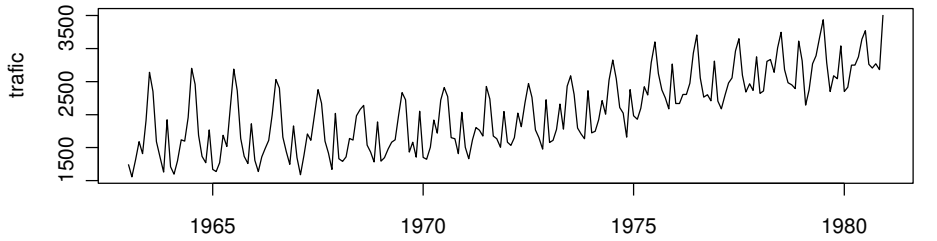


Fig.5.11–TraficdunombredevoyageursSNCF**EnlangageR**trafic=c(1750,1560,1820,2090,1910,2410,3140,285

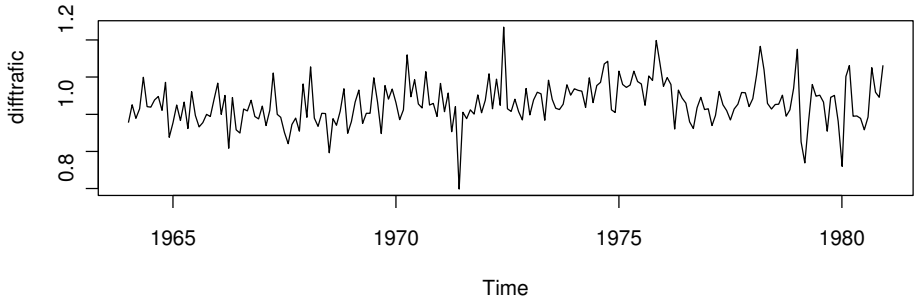


Fig.5.12–Diff´erenced'ordre12surlasérietraficdunombredevoyageursSNCF

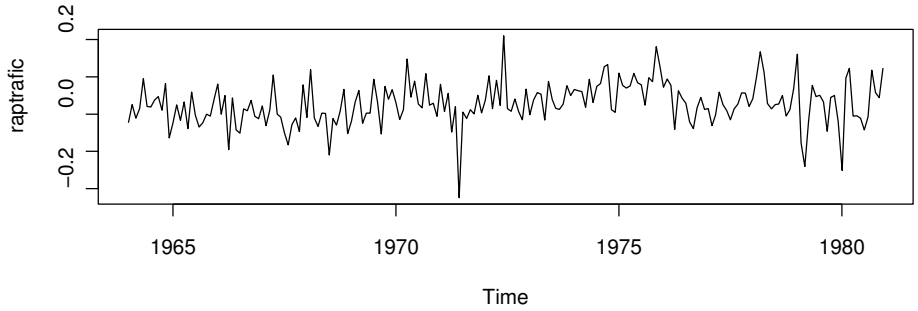


Fig.5.13–Logarithmedurapportd'ordre12surlas´erietraficdunombredevoyageursSNCF

1834,1792,1860,2138,2115,2485,2581,2639,2038,1936,1784,2391,1798,1850,1981,2085,2120,2491,2834
806,2976,3430,3705,3053,2764,2802,2707,3307,2706,2586,2796,2978,3053,3463,3649,3095,2839,2966,
´eaires Unfiltrelin´eaired'ordre$m=p_1+p_2$estd´efinipar$FL=_{p_2 j=-p_1} w_j L^{-j}=w_{-p_1} L^{p_1}+w_{-p_1+1} L^{p_1-1}+\cdots+w_{-1} L+w_0 I+$

o`u$p_1, p_2 \in \mathbb{N}$et$w_j \in \mathbb{R}$.

5.4.2Moyennesmobiles:définition

Unemoyennemobiled'ordre$m = p_1 + p_2 + 1$estunfiltrelin´eairetelque

$$\sum_{j=-p_1}^{p_2} w_j = 1, \text{ pourtout} j = -p_1, \ldots, p_2.$$

Beaucoupdemoyennesmobilesontdespoids$w_j$positifs,maispastoutes.

Unemoyennemobileestsym´etriquesi$p_1 = p_2 = p$,et

$$w_j = w_{-j}, \text{ pourtout} j = 1, \ldots, p.$$

Unemoyennemobilesym´etriqueestditenon-pond´er´eesi

$$w_j = cst \text{ pourtout} j = -p_1, \ldots, p_2.$$

5.4.3Moyennemobileetcomposantesaisonnière

Unemoyennemobileestunoutilint´eressantpourlisserunes´erietemporelleetdoncpourenleverune composantesaisonni`ere.Onutilisedepr´ef´erencedesmoyennesmobilesnon-pond´er´eesd'ordreégal`ala p´eriode,parexemped'ordre7pourdesdonn´eesjournali`eres,d'ordre12pourdesdonn´eesmensuelles.Par exemple,pourenleverlacomposantesaisonni`eredueaujourdelasemaine,onpeutappliquerunemoyenne mobilenon-pond´er´eed'ordre7.$MM(7) = \frac{1}{7}(L_3 + L_2 + L + I + F + F_2 + F_3)$

.

Cettemoyennemobileaccordelemˆemepoids`achaquejourdelasemaine.Eneffet,

$MM(7)y_t = \frac{1}{7}(y_{t-3} + y_{t-2} + y_{t-1} + y_t + y_{t+1} + y_{t+2} + y_{t+3})$.Pourlescomposantessaisonni`eresd'unep´eriodepaire,il

ees

non-pond´er´ees.Ilexistedeuxtypesdemoyennemobilecentr´ee:–Silap´eriodeestpaireet´egale`am,($m = 4$pour

mobiled'ordreimpairaccordantundemi-poidsauxdeuxextr´emit´es.Parexemple,pourdesdonn´ees trimestrielles,lamoyennemobileestd´efinieparMM(4)$= \frac{1}{8}(L_2 + 2L + 2I + 2F + F_2)$.Ainsi,chaquetrimestreconservelen ´eesseront"perdues"auxextr´emit´esdelas´eries.**Exemple5.8**Lavariable"r´efrig´erateur"estliss´eegrˆace`a

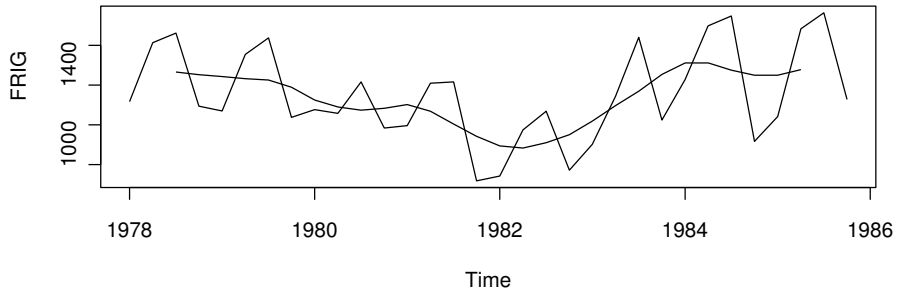


Fig.5.14–Nombredeérateursetmoyennemobiled'ordre4

EnlangageRdec=decompose(FRIG)moving_average=dec$trendplot(FRIG)lines(moving_average)Unemoyen

sonni`ere.**5.5Moyennesmobilesparticuli`eres5.5.1MoyennemobiledeVanHan**
´eriode4et5etconservelestendancespolynomialesjusqu'`al'ordre3.**5.5.3MoyennemobiledeHender**

Moyennemobiledе Hendersond'ordre $2m-3$, **où** $m \geq 4$

$$\mathrm{MM}_H = \sum_{j=-m-1}^{m+1} \theta_j L^j,$$

où $\theta_j =$

$$\frac{315((m-1)^2 - i^2)(m^2 - i^2)((m+1)^2 - i^2)(3m^2 - 16 - 11i^2)}{8m(m^2-1)(4m^2-1)(4m^2-9)(4m^2-25)}$$

Moyennemobiledе Hendersond'ordre $2m-3 = 5 (m=4)$

$$\frac{1}{286}(-21L^2 + 84L + 160I + 84F - 21F^2)$$

Moyennemobiledе Hendersond'ordre $2m-3 = 9 (m=6)$

$$\frac{1}{24310}(-99L^4 - 24L^3 - 288L^2 + 648L + 805I + 648F + 288F^2 - 24F^3 - 99F^4)$$

Moyennemobiledе Hendersond'ordre $2m-3 = 11 (m=7)$

$$\frac{1}{923780}(-2574L^5 - 2475L^4 + 3300L^3 + 13050L^2 + 22050L + 25676I$$

$$+ 22050F + 13050F^2 + 3300F^3 - 2475F^4 - 2574F^5)$$

Moyennemobiledе Hendersond'ordre $2m-3 = 15 (m=9)$ $1193154(-2652L^7 - 4732L^6 - 2730L^5 + 4641L^4 + 160$

$$+ 37422F + 28182F^2 + 16016F^3 + 4641F^4 - 2730F^5 - 4732F^6 - 2652F^7)$$

5.5.4 M´edianesmobiles

Silesdonn´eescontiennentdesvaleursaberrantesouextrˆemes, onpeutremplac

unem´edianemobile. Parexemplelam´edianemobiled'ordre 5 estd´efiniepar: $Med(5)_t = \mathrm{M´ediane}(y_{t-2}, y_{t-1}, y_t, y_{t+}$
´eesetlatendance $S_m = 1A - 1_a(Y_{am} - T_{am})$. 72

Eng´eneral,onnedisposepasdunenombred'observations,pourchaquemois.Onprocèdeàunajustement afinquelasommedescomposantessaisonnièressoitégale`azéro:

$$S_m = S_m - \frac{1}{M} \sum_m S_m.$$

Onpeutensuiteproc´eder`aladésaisonnalisationdelas´erie par

$$Y_{am} = Y_{am} - S_m.$$

5.6.2M´ethodemultiplicative

Soitune´serietemporelle´egieparunmod`elemultiplicatifdutype

$$Y_{am} = T_{am} \times S_m \times E_{am}.$$

o`u$a = 1,\ldots,A$repr´esenteparexemplel'ann´eeet$m = 1,..,M$repr´esenteparexemplelemois.Latendance estsuppos´eeconnuesoitparunajustement,soitparunemoyennemobile.

Onisolelacomposantesaisonni`ereenfaisant,pourchaquemois,lamoyennedesrapports entrelesvaleurs observ´eesetlatendance:$S_m = 1A - 1_a Y$

$$\frac{am}{T_{am}}.$$

`Anouveau,onr´ealiseunajustementafinquelamoyennedescomposantessaisonni`eressoitégale`a1.On corrigedonclescoefficients$S_m$par$S_m = S_m 1 1_M m S_m$.Lad´esaisonnalisationser´ealisealorsparunedivision$Y_{am} = Y_{am} S$

erateursLeTableau5.4contientlavariable'ventedeer´efrig´erateurs',lamoyennemobiled'ordre4,lacomposantedes´esaisonni` ets´eried´esaisonnalis´eeaumoyendelam´ethodeadditive.LeTableau5.6pr´esentelad´esaisonnalisationau moyendelam´ethodemultiplicative.**EnlangageR**deco=decompose(FRIG,type="multiplicative")plot(deco)**5.** ´ediction`al'horizon1,etconsiste`ar´ealiserunemoyennedesvaleurspass´eesenaffectantdespoidsmoinsimpor

Tab.5.4–D´ecompositiondelavariable FRIG, m´ethodeadditive

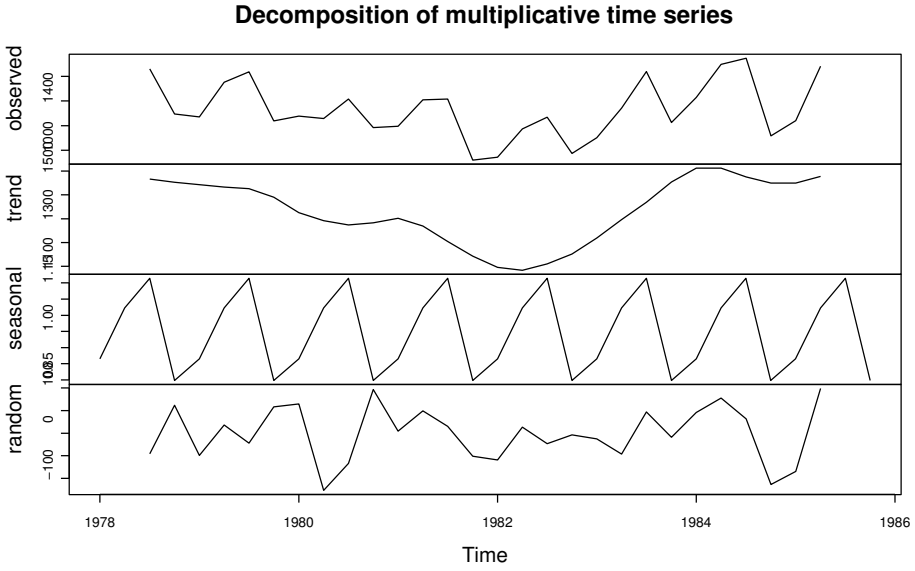
QTR	FRIG	MM	FRIG-MM	Desaison
1	1317			1442.58
2	1615			1505.13
3	1662	1466.50	195.50	1451.20
4	1295	1453.25	-158.25	1490.09
5	1271	1442.88	-171.88	1396.58
6	1555	1432.88	122.13	1445.13
7	1639	1426.50	212.50	1428.20
8	1238	1390.13	-152.13	1433.09
9	1277	1325.25	-48.25	1402.58
10	1258	1290.88	-32.88	1148.13
11	1417	1274.13	142.88	1206.20
12	1185	1283.00	-98.00	1380.09
13	1196	1302.00	-106.00	1321.58
14	1410	1268.75	141.25	1300.13
15	1417	1203.88	213.13	1206.20
16	919	1142.88	-223.88	1114.09
17	943	1095.00	-152.00	1068.58
18	1175	1083.25	91.75	1065.13
19	1269	1109.88	159.13	1058.20
20	973	1150.88	-177.88	1168.09
21	1102	1218.50	-116.50	1227.58
22	1344	1296.50	47.50	1234.13
23	1641	1368.88	272.13	1430.20
24	1251	1454.13	-229.13	1420.09
25	1429	1512.00	-83.00	1554.58
26	1699	1512.00	187.00	1589.13
27	1749	1475.1		

Tab.5.6–D´ ecompositiondelavariableFRIG,méthodemultiplicative

QTR	FRIG	MM	FRIG/MM	Desaison
1	1317			1453.85
2	1615			1493.76
3	1662	1466.50	1.13	1434.00
4	1295	1453.25	0.89	1516.45
5	1271	1442.88	0.88	1403.07
6	1555	1432.88	1.09	1438.26
7	1639	1426.50	1.15	1414.15
8	1238	1390.13	0.89	1449.70
9	1277	1325.25	0.96	1409.70
10	1258	1290.88	0.97	1163.56
11	1471	1274.13	1.11	1222.61
12	1185	1283.00	0.92	1387.64
13	1196	1302.00	0.92	1320.28
14	1410	1268.75	1.11	1304.15
15	1417	1203.88	1.18	1222.61
16	919	1142.88	0.80	1076.15
17	943	1095.00	0.86	1040.99
18	1175	1083.25	1.08	1086.79
19	1269	1109.88	1.14	1094.91
20	973	1150.88	0.85	1139.39
21	1102	1218.50	0.90	1216.51
22	1344	1296.50	1.04	1243.10
23	1641	1368.88	1.20	1415.88
24	1225	1454.13	0.84	1434.48
25	2514	2912.00	0.95	1577.49
26	1699	1512.00	1.12	1571.45
27	1747	1491.47	5.13	1.19

´eeparunedroitehorizontale.Autrementdit,onsupposeque$X_T \approx a.75$

Fig.5.15 – D´ ecompositiondelas´eriedeventesderéfrigérateurs5.1



Lelissageexponentielpeutˆetreobtenuaumoyendelam´ethodedesmoindres carrés en minimisantena lecrit`ere$Q=\sum_{j=0}^{T-1}\beta_j(X_{T-j}-a)^2$.Enannulantlad´eriv´eeparrapport`aa,onobtient$2\sum_{j=0}^{T-1}\beta_j(X_{T-j}-a)=0$,cequidonne ´eeparunedroitequelconquedetype$a+b(t-T)$.Onappliquealorsunlissageexponentieldoublepourobtenirlapr´ec

Enannulantlesdérivéespartiellesparrapport`aaet$b$,onobtient

$$\begin{cases} 2\sum_{j=0}^{T-1}\beta^j\,(X_{T-j}-a+bj)=0 \\ 2\sum_{j=0}^{T-1}\beta^j\,(X_{T-j}-a+bj)j=0. \end{cases}$$

cequidonne

$$\begin{cases} \sum_{j=0}^{T-1}\beta^j X_{T-j}-a\sum_{j=0}^{T-1}\beta^j+b\sum_{j=0}^{T-1}j\beta^j=0 \\ \sum_{j=0}^{T-1}j\beta_j X_{T-j}-a\sum_{j=0}^{T-1}j\beta^j+b\sum_{j=0}^{T-1}j^2\beta^j=0. \end{cases}$$

Commeona∞

$$\sum_{j=0}\beta^j=\frac{1}{1-\beta}$$

$$\sum_{j=0}j\beta_j=\beta\frac{}{(1-\beta)^2}$$

$\sum_{j=0}j^2\beta_j=\beta(1+\beta)(1-\beta)^3ona\sum_{j=0}^{T-1}\beta_j X_{T-j}-a1-\beta+b\beta(1-\beta)^2=0\sum_{j=0}^{T-1}j\beta_j X_{T-j}-a\beta(1-\beta)^2+b\beta(1+\beta)(1-\beta$

EnnotantmaintenantS_{1T}las´erieliss´ee$S_{1T}=(1-\beta)\sum_{j=0}^{T-1}\beta_j X_{T-j}$,et$S_{2T}$las´eriedoublementliss´ee$S_{2T}=(1-\beta)\sum_{j=0}^{T-1}\beta$

Lesystème(5.1)peutalorsse ́crire

$$\begin{cases} \dfrac{S_T^1}{1-\beta} - \dfrac{a}{1-\beta} + \dfrac{b\beta}{(1-\beta)^2} = 0 \\ \dfrac{S_T^2}{(1-\beta)^2} - \dfrac{S_T^1}{1-\beta} - \dfrac{a\beta}{(1-\beta)^2} + \dfrac{b\beta(1+\beta)}{(1-\beta)^3} = 0. \end{cases}$$

Enr ́esolvantcesyst`emeen$a$et$b$,onobtientfinalement

$$\begin{cases} a = 2S_T^1 - S_T^2 \\ b = \dfrac{1-\beta}{\beta}(S_T^1 - S_T^2). \end{cases}$$

Exemple5.10 Letableau5.8rendcompteduprixmoyendumazoutpour100 (achatentre800et1500)enCHFpourchaquemoisde2004`a2007(Source:Officefe ́de ́raldelastatistique,2008).

Tab.5.8–PrixmoyenduMazoutpour100 (achatentre800et1500)

mois/ann ́ee	2004	2005	2006	2007
janvier	54.23	63.00	86.16	79.39
f ́evrier	51.51	67.32	88.70	81.32
mars	55.60	75.52	88.92	82.06
avril	55.72	79.83	92.58	88.05
mai	58.71	73.22	93.65	88.24
juin	58.82	75.38	91.88	88.95
juillet	58.41	83.97	95.35	92.10
ao ̂ut	64.92	84.23	95.83	9

 ́ecursive $\hat{X}_t(1) = (1-\beta)X_t + \beta\hat{X}_{t-1}(1)$, $\hat{X}_0(1) = X_1$, o`u $S_{1t} = \hat{X}_t(1)$. Onobtient: $S_{11} = \hat{X}_1(1) = (1-\beta)X_1 + \beta\hat{X}_0(1) = (1$

Onobtient:

$$S_1^2 = (1-\beta)S_1^1 + \beta S_0^2 = (1-\beta)S_1^1 + \beta S_1^1 = S_1^1 = 54.23,$$
$$S_2^2 = (1-\beta)S_2^1 + \beta S_1^2 = 0.3\times 53.414 + 0.7\times 54.23 = 53.99,$$
$$S_3^2 = (1-\beta)S_3^1 + \beta S_2^2 = 0.3\times 54.070 + 0.7\times 53.99 = 54.01,$$

etainsidesuite.–Oncherchealorsˆ

$$X_t(k) = a + bk$$

pourchaque$t$.Onprendi$c$ $k=1$, $\hat{X}_t(1) = a+b$avec:

$$a = 2S_t^1 - S_t^2$$
$$b = \frac{1-\beta}{\beta}\left(S_t^1 - S_t^2\right) = \frac{0.3}{0.7}\left(S_t^1 - S_t^2\right)$$

Letableau5.9rendcomptedesr´esultatspourlesann´ees2004`a2007.

Lafigure5.16repr´esentelas´erieinitiale,lelissageexponentielsimpleetlelissageexponentieldoubleet peutˆetreobtenueenlanguageRaumoyenducodesuivant:

```
#Lissageexponentieldoubleaveck=1mazout=c(54.23,51.51,55.60,55.72,58.71,58.82,+58.41,64.92,63.95,7

+79.83,73.22,75.38,83.97,84.23,97.29,99.31,89.88,87.18,+86.16,88.70,88.92,92.58,93.65,91.88,95.35,95.
+p*liss2[i-1]#formuler´ecursive}#Lissageexponentieldoubleaveck=1a=2*liss-liss2b=((1-p)/p)*(liss-liss2)lis
```

Tab.5.9–Lissageexponentielsimpleetdoubledelasérietemporelle Prixmoyendu Mazoutpour100litres (achatentre800et1500litres)enCHF

année	mois	X_t	S_t^1	S_t^2	a	b	$a+b$
2004	1	54.23	54.230	54.230	54.230	0.000	54.230
	2	51.51	53.414	53.985	52.843	-0.245	52.598
	3	55.60	54.070	54.011	54.129	0.025	54.154
	4	55.72	54.564	54.177	54.952	0.166	55.119
	5	58.71	55.808	54.666	56.950	0.489	57.440
	6	58.82	56.712	55.280	58.144	0.614	58.757
	7	58.41	57.221	55.862	58.580	0.582	59.163
	8	64.92	59.531	56.963	62.099	1.101	63.199
	9	63.95	60.857	58.131	63.582	1.168	64.750
	10	72.98	64.494	60.040	68.947	1.909	70.856
	11	70.25	66.221	61.894	70.547	1.854	72.401
	12	68.24	66.826	63.374	70.279	1.480	71.759
2005	1	63.00	65.678	64.065	67.292	0.691	67.983
	2	67.32	66.171	64.697	67.645	0.632	68.277
	3	75.52	68.976	65.981	71.971	1.284	73.254
	4	79.83	72.232	67.856	76.608	1.875	78.483
	5	73.22	72.528	69.256	75.799	1.402	77.201
	6	75.38	73.384	70.496	76.272	1.238	77.510
	7	83.97	76.560	72.315	80.805	1.819	82.624
	8	84.23	78.861	74.279	83.443	1.964	85.407
	9	97.29	84.390	77.312	91.467	3.033	94.501
	10	99.31	88.867	80.778	96.953	3.466	100.420
	11	89.88	89.170	83.296	95.044	2.518	97.562
	12	87.18	88.573	84.879	92.267	1.583	93.850

2006186.1687.84985.77089.9280.89190.819288.7088.10486.47089.7380.70090.439388.9
1.41380.486281.3284.03487.15680.911-1.33879.573382.0683.44186.04180.842-1.11479

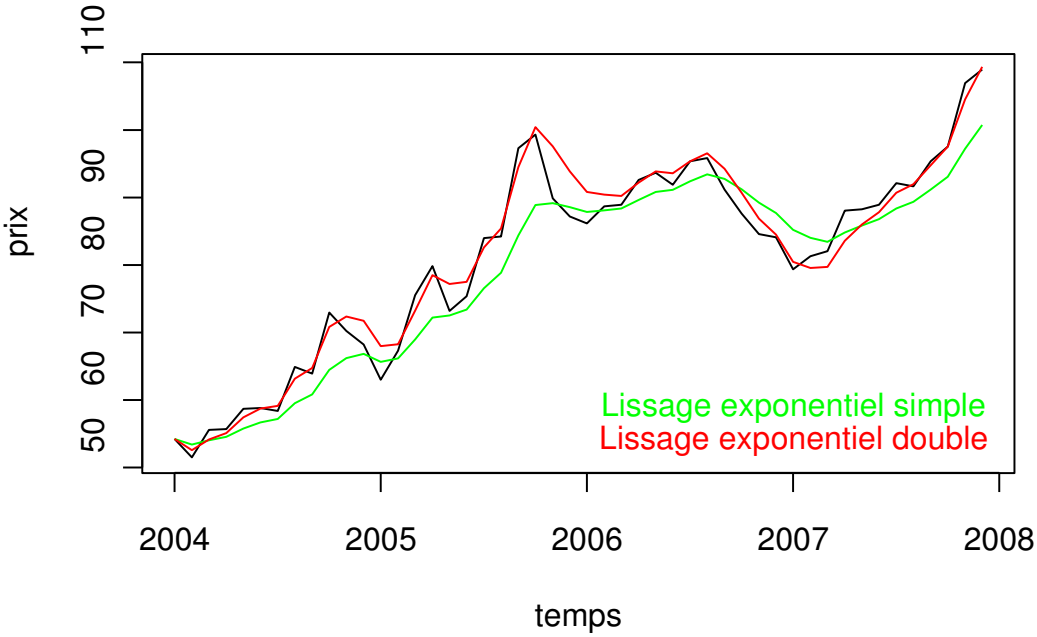


Fig.5.16–EvolutionduprixdumazoutenCHF(achatentre800et1500),lissageexponentieldoubleetlissageexponen

Exercices

Exercice5.1 Désaisonnalisezlasérie suivante(c'estunesérietrimestriellesur3années)

2417,1605,1221,1826,2367,1569,1176,1742,2804,1399,1063,1755

parlaméthodeadditive,enutilisantunemoyennemobiled'ordre4.

Exercice5.2 EnlangageRutilisezlasérie"ldeaths"quiestunesérieexistetrouvedanslepackagedebase "datasets".Lisezladocumentation,puisdésaisonnalisezcetteséparlesméthodesadditiveetmultiplicative.82

Chapitre 6

Calcul des probabilités et variables aléatoires

6.1 Probabilités

6.1.1 Événement

Une expérience est dite aléatoire si on ne peut pas prévoir avec certitude le résultat. On note ω un résultat possible de cette expérience aléatoire. L'ensemble de tous les résultats possibles est appelé l'univers et est noté Ω. Par exemple, si on jette deux pièces de monnaie, on peut obtenir les résultats $\Omega = \{(P,P),(F,P),(P,F),(F,F)\}$, avec F pour "face" et P pour "pile".

Un événement est un sous-ensemble de Ω. Par exemple, l'événement "avoir deux fois pile" ou "avoir au moins une fois pile". Formellement, un événement est un sous-ensemble de Ω. – L'événement "avoir deux fois pile" est le sous ensemble $\{(P,P)\}$. – L'événement "avoir au moins

L'ensemble Ω est appelé événement certain, et l'ensemble vide $\emptyset$ est appelé événement impossible.

6.1.2 Opérations sur les événements

Sur les événements, on peut appliquer les opérations habit
événement A est "obtenir un nombre pair" et l'événement B "obtenir un multiple de 3", l'événement $A \cap B$ est l'événe

—

Exemple6.1 L'expériencepeutconsister`ajeterundé,alors

$$\Omega = \{1,2,3,4,5,6\},$$

etun´ev´enement,notéA,est"obtenirunnombrepair".Onaalors

$$A = \{2,4,6\} \text{ et } \overline{A} = \{1,3,5\}.$$

6.1.3 Relationsentrelesévénements

´Ev´enementsmutuellementexclusifs

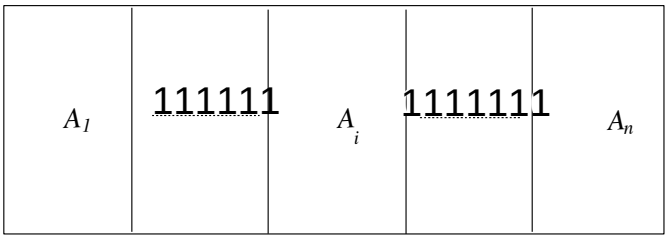
Si$A \cap B = \emptyset$onditque$A$et$B$sontmutuellementexclusifs,cequisignifieque$A$et$B$nepeuventpas seproduireensemble.**Exemple6.2** Sionjetteund´e,l'´ev´enement"obtenirunnombrepair"etl'´ evénement"obtenirunnombre impair"nepeuventpasˆetreobtenusenmˆemetemps.Ilssontmutuellementexclusifs.D'autrepart,si l'on jetteund´e,les´ev´enements$A$:"obtenirunnombrepair"n'estpasmutuellementexclusifavecl'événement B:"obtenirunnombreinf´erieurou´egal`a3".Eneffet,l'intersectionde$A$et$B$estnon-videetconsisteen l'´ev´enement"obtenir2".**Inclusion** SiAestinclusdans$B$,on´ecrit$A \subset B$.Onditque$A$implique$B$.**Exemple6.3** Sionje

$A = \{2\}$et$B = \{2,4,6\}$.Onditque$A$implique$B$.

6.1.4 Ensembledespartiesd'unensembleetsyst`eme

´e$P(.)$estuneapplicationde$A$dans$[0,1]$,*telleque:*–$\Pr(\Omega) = 1$,–*Pourtoutensembled´enombrabled'´ev´enements*A

Tab.6.1–Syst`emecompletd'ev´enements



Apartirdesaxiomes,onpeutd´eduirelespropri´et´essuivantes:

Propri´et´e6.1 Pr(∅)=0.*D´emonstration* Comme∅estd'intersectionvideavec∅,onaquePr(∅∪∅)=Pr(∅)+Pr(∅

Donc,Pr(∅)=2Pr(∅),cequiimpliquequePr(∅)=0.✷

Propri´et´e6.2 $\Pr(\overline{A})=1-\Pr(A)$.*D´emonstration* Onsaitque$A\cup A=\Omega$et$A\cap A=\emptyset$.Ainsi,onaquePr(Ω)=Pr(*A*∪*A*)=Pr(*A*

Propri´et´e6.3 Pr(*A*)≤Pr(*B*)*si*$A\subset B$.*D´emonstration* Comme$A\subset B$,ona$B=(B\cap A)\cup A$.Maisonaque$(B\cap A)\cap A=\emptyset$.Ains

D´emonstration
On aque $A \cup B = A \cup (B \cap A)$,

avec $A \cap (B \cap A = \emptyset)$.

Ainsi $Pr(A \cup B) = Pr(A) + Pr(B \cap A)$.

Il reste à montrer que $Pr(B) = Pr(B \cap A) + Pr(A \cap B)$

Mais $B = (B \cap A) \cup (B \cap A)$

avec $(B \cap A) \cap (B \cap A) = \emptyset$

Donc $Pr(B) = Pr(B \cap A) + Pr(B \cap A)$

Propri´et´e 6.5 $Pr_{ni=1} A_i \leq_{ni=1} Pr(A_i)$ *D´emonstration* Notons respectivement $B_1 = A_1, B_2 = (A_2 \setminus A_1), B_3 = (A_3 \setminus (A_1 \cup A$

$_2$)),

$B_4 = (A_4 \setminus (A_1 \cup A_2 \cup A_3)), \ldots, B_n = (A_n \setminus (A_1 \cup A_2 \cup A$ $_3 \cup \cdots \cup A$ $_{n-1}$)).

Comme $_{ni=1} A_i =_{ni=1} B_i$, et que $B_i \cap B_j$ pour tout $j = i$, alors $Pr_{ni=1} B_i =_{ni=1} Pr(B_i)$. De plus, comme, pour tout $i, B_i \subset A_i$, on a que $Pr($

6.1.6 Probabilités conditionnelles et indépendance

Définition 6.3 *Soient deux événements A et B, si* $\Pr(B)>0$, *alors*

$$\Pr(A|B)=\frac{\Pr(A\cap B)}{\Pr(B)}.$$

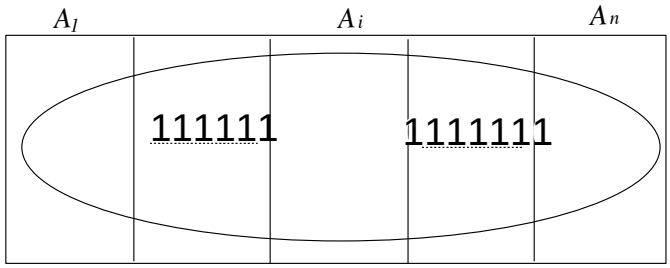
Exemple 6.5 Sion jette un dé, et quel'on considère les deux événements suivants:
– A la probabilité d'avoir un nombre pair et
– B la probabilité d'avoir un nombre supérieur ou égal à 4.
On a donc – $\Pr(A)=\Pr(\{2,4,6\})=12$, – $\Pr(B)=\Pr(\{4,5,6\})=36=12$, – $\Pr(A\cap B)=\Pr(\{4,6\})=26=13$, – $\Pr(A|B)=\Pr(A\cap B)\Pr$

$\Pr(A|B)=\Pr(A)$. On peut montrer facilement que si A et B sont indépendants, alors $\Pr(A\cap B)=\Pr(A)\Pr(B)$.

6.1.7 Théo

$\Pr(B)=\sum_{i=1}^{n}\Pr(A_i)\Pr(B|A_i)$.

Tab. 6.2 – Illustration du théorème des probabilités totales



En effet, $\sum_{i=1}^{n}\Pr(A_i)\Pr(B|A_i)=\sum_{i=1}^{n}\Pr(B\cap A_i)$. Comme les événements $A_i\cap B$ sont mutuellement exclusifs, $\sum_{i=1}^{n}\Pr(B\cap A_i)=$

**Th´eor`eme6.2** *(deBayes)SoitA* $_1$*,...,A* $_n$ *unsyst`emecompletd'´ev´enements,alors*

$$\Pr(A_i|B) = \frac{\Pr(A_i)\Pr(B|A_i)}{\sum_{j=1}^{n}\Pr(A_j)\Pr(B|A_j)}.$$

Eneffet,parleth´eor`emedesprobabilit´estotales,

$$\frac{\Pr(A_i)\Pr(B|A_i)}{\sum_{j=1}^{n}\Pr(A_j)\Pr(B|A_j)} = \frac{\Pr(B\cap A_i)}{\Pr(B)} = \Pr(A_i|B).$$

Exemple6.6 Supposonsqu'unepopulationd'adultessoitcompos´eede30%defumeurs(A_1)etde70%denon-fumeur(A_2).Notons$B$l'´ev´enement"mourird'uncancerdupoumon"Supposonsenoutreque laprobabilit´edemourird'uncancerdupoumonest´egale`aPr($B|A_1$) =20%si l'onestfumeuretde Pr($B|A_2$)=1%sil'onestnon-fumeur.Leth´eor`emedeBayespermetdecalculerlesprobabilit´esapriori, c'est-`a-direlaprobabilit´ed'avoir´et´efumeursionestmortd'uncancerdupoumon.Eneffet,cetteprobabilit´ estnot´eePr($A_1|B$)etpeut^etrecalcul´eeparPr($A_1|B$)=Pr(A_1)Pr($B|A_1$)Pr(A_1)Pr($B|A_1$)+Pr(A_2)Pr($B|A_2$)=

$$\frac{0.3\times 0.2}{0.3\times 0.2+0.7\times 0.01} = \frac{0.06}{0.06+0.007} \approx 0.896.$$

Laprobabilit´edenepasavoir´et´enonfumeursionestmortd'uncancerdupoumonvautquant`aelle:

$$\Pr(A_2|B) = \Pr(A_2)\Pr(B|A_2)\Pr(A_1)\Pr(B|A_1)+\Pr(A_2)\Pr(B|A_2) = \frac{0.7\times 0.01}{0.3\times 0.2+0.7\times 0.01} = \frac{0.07}{0.06+0.007} \approx 0.104.$$

6.2Analysecombinatoire

6.2.1Introduction

L'analysecombinatoireestl'´etudemath´emat

natoireestunoutilutilis´edanslecalculdesprobabilit´es.

6.2.2Permutations(sansr´ep´etition)

Unep

l'ensemble$\{1,2,3\}$.Ilexiste6mani`eresd'ordonnercestroischiffres:$\{1,2,3\},\{1,3,2\},\{2,1,3\},\{2,3,1\},\{3,1,2\},\{3$

6.2.3 Permutations avec répétition

On peut également se poser la question du nombre de manières de ranger des objets qui ne sont pas tous distincts. Supposons que nous ayons 2 boules rouges (notées R) et 3 boules blanches (notées B). Il existe 10 permutations possibles qui sont :

$$\{R,R,B,B,B\},\{R,B,R,B,B\},\{R,B,B,R,B\},\{R,B,B,B,R\},\{B,R,R,B,B\},$$

$$\{B,R,B,R,B\},\{B,R,B,B,R\},\{B,B,R,R,B\},\{B,B,R,B,R\},\{B,B,B,R,R\}.$$

S'il y a n objets appartenant à deux groupes de tailles n_1 et n_2, le nombre de permutations avec répétitions est $n!$

$$\overline{n_1!n_2!}.$$

Par exemple s'il y a 3 boules blanches et 2 boules rouges, on obtient

$$\overline{n!n_1!n_2!} = \frac{5!}{2!3!} = \frac{120}{2\times 6} = 10.$$

S'il y a n objets appartenant à p groupes de tailles $n_1, n_2, \ldots, n_p$, le nombre de permutations avec répétitions est $n!n_1!n_2!\times\cdots\times n_p!$.

6.2.4 Arrangements (sans répétition)

Soit n objets distincts. On appelle un arrangement une manière de sélectionner k objets parmi les n et de les ranger dans des boîtes numérotées de 1 à k. Dans la première boîte, on peut mettre chacun des n objets. Dans la deuxième boîte, on peut mettre chacun des $n-1$ objets restants, dans la troisième boîte, on peut mettre chacun des $n-2$ objets restants et ainsi de suite. Le nombre d'arrangements possibles est donc égal à : $A_k^n = n\times(n-1)\times(n-2)\times\cdots\times(n-k+1) = n!(n-k)$

---

tenir compte de leur ordre. Le nombre de combinaisons est le nombre de sous-ensembles de taille k dans un ensemble de taille n. Soit l'ensemble $\{1,2,3,4,5\}$. Il existe 10 sous-ensembles de taille 3 qui sont : $\{1,2,3\},\{1,2,4\},\{1$
ements, s'il on ne veut pas tenir compte de l'ordre on divise A_k^n par $k!$. Le nombre de combinaisons de k objets parmi n va

6.3Variablesal´eatoires

6.3.1D´efinition

Lanotiondevariableal´eatoireformalisel'associationd'unevaleurau r´esultatd'uneexp´erienceal´eatoire.

D´efinition6.5 *Unevariableal´eatoire*X*estuneapplicationdel'ensemblefondamental*Ω*dans*$\mathbb{R}$.

Exemple6.7 Onconsid`ereuneexp´erienceal´eatoireconsistant`alancerdeuxfois demonnaie.L'ensemble desr´esultatspossiblesest$\Omega = \{(F,F),(F,P),(P,F),(P,P)\}$.

Chacundes´el´ementsdeΩauneprobabilit´e1/4.Unevariablealéatoirevaassocierunevaleur`achacundes ´el´ementsdeΩ.Consid´eronslavariableal´eatoirerepr´esentantlenombrede"Faces"obtenus:

$$X = \begin{cases} 0 & \text{avecuneprobabilit´e1/4} \\ 1 & \text{avecuneprobabilit´e1/2} \\ 2 & \text{avecuneprobabilit´e1/4.} \end{cases}$$

C'estunevariableal´eatoirediscr`etedontladistributiondeprobabilit´e estpr´esent´eeenFigure6.1.

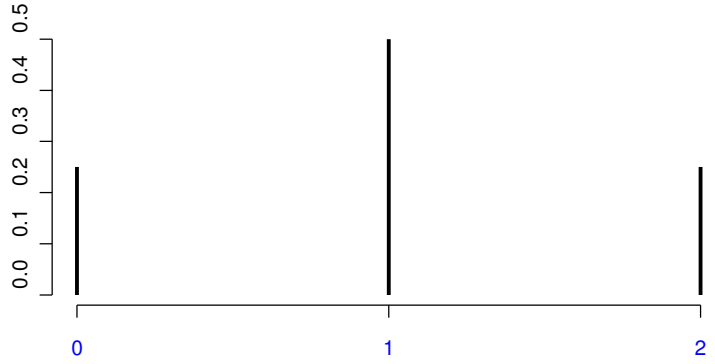


Fig.6.1–Distributionde"faces"obtenus.

6.4Variablesal´eatoiresdiscr`etes

6.4.1D´efini

$(x-\mu)^2 = \sum_{x \in Z} p_X(x) x^2 - \mu^2$.Onpeutaussicalculerlesmomentsettouslesautresparam`etres.

6.4.2 Variable indicatrice ou bernoullienne

La variable indicatrice X de paramètre $p \in [0,1]$ a la distribution de probabilités suivante:

$$X = \begin{array}{l} 1 \text{ avec une probabilité } p \\ 0 \text{ avec une probabilité } 1-p. \end{array}$$

L'espérance vaut $\mu = \mathrm{E}(X) = 0 \times (1-p) + 1 \times p = p,$

et la variance vaut $\sigma^2 = \mathrm{var}(X) = \mathrm{E}(X-p)^2 = (1-p)(0-p)^2 + p(1-p)^2 = p(1-p).$

Exemple 6.8 On tire au hasard une boule dans une urne contenant 18 boules rouges et 12 boules blanches. Si X vaut 1 si la boule est rouge et 0 sinon, alors X a une loi bernoullienne de paramètre $p = 18/(18+12) = 0.6.$

6.4.3 Variable binomiale

La variable aléatoire binomiale de paramètres n et p correspond à l'expérience suivante. On renouvelle n fois de manière indépendante une épreuve de Bernoulli de paramètre p, où p est la probabilité de succès pour une expérience élémentaire. Ensuite, on note X le nombre de succès obtenus. Le nombre de succès est une variable aléatoire prenant des valeurs entières de 0 à n et ayant une distribution binomiale.

Une variable X suit une loi binomiale de paramètre $0 < p < 1$ et d'exposant n, si

$$\Pr(X = x) = n_x p^x q^{n-x}, \quad x = 0, 1, \ldots, n-1, n,$$

où $q = 1-p$, et $n_x = n!/x!(n-x)!$. De manière synthétique, si X a une distribution binomiale, on note: $X \sim B(n,p)$. **Rappe**

---

du binôme de Newton $(p+q)^n$. $(p+q)^0 = 1$ $(p+q)^1 = p+q = 1$ $(p+q)^2 = p^2 + 2pq + q^2 = 1$ $(p+q)^3 = p^3 + 3p^2q + 3pq^2 + q^3 = 1$ (

L'espérancesecalculedelamanièresuivante:

$$
\begin{aligned}
\mathrm{E}(X) &= \sum_{x=0}^{n} x \Pr(X=x) \\
&= \sum_{x=0}^{n} x \binom{n}{x} p^x q^{n-x} \\
&= \sum_{x=1}^{n} x \binom{n}{x} p^x q^{n-x} \quad \text{(onpeutenleverleterme } x=0\text{)} \\
&= n \sum_{x=1}^{n} \binom{n-1}{x-1} p^x q^{n-x} \\
&= np \sum_{x=1}^{n} \binom{n-1}{x-1} p^{x-1} q^{(n-1)-(x-1)} \\
&= np \sum_{z=0}^{n-1} \binom{n-1}{z} p^z q^{(n-1)-z} \quad \text{(enposant } z=x-1\text{)}
\end{aligned}
$$

$=np(p+q)^{n-1}=np$.Lavarianceestdonn´ee(sansd´emonstration)parvar$(X)=npq$.**Exemple6.9**Ontireauhasarda

18boulesrougeset12boulesblanches.SiXestlenombredeboulesrougesobtenues,alors$X$auneloi
binomialedeparam`etre$p=18/(18+12)=0.6$,etd'exposant$n=5$.Donc,Pr$(X=x)=5^x0.6^x0.4^{5-x}, x=0,1,...,4,5$,cequ
´eedanslaFigure6.2.**Exemple6.10**Supposonsque,dansunepopulationd'´electeurs,60%des´electeurss'appr^e

_____

_____

_____

_____

_____

_____

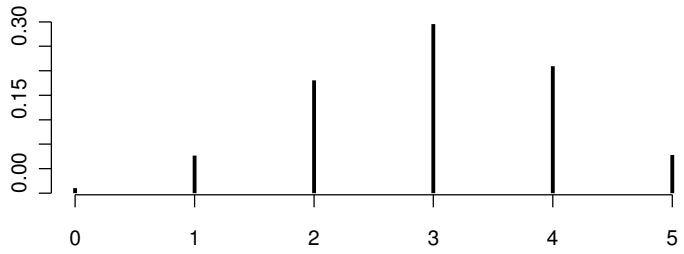


Fig.6.2–Distributiond'unevariableleat´oirebinomialeavec$n$=5et$p$=0.6.

6.4.4VariabledePoisson Lavariable$X$suituneloidePoisson,deparam`

etre$\lambda \in \mathbb{R}^+$ si

$$\Pr(X=x)=e^{-\lambda}\frac{\lambda^x}{x!}, x=0,1,2,3,.....$$

Onnotealors$X \sim P(1)$.Lasommedesprobabilit´esestbien´egale`a1,eneffet

$$\sum_{x=0}^{\infty}\Pr(X=x)=\sum_{x=0}^{\infty}e^{-\lambda}\frac{\lambda^x}{x!}=e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^x}{x!}=e^{-\lambda}e^{\lambda}=1.$$

L'esp´eranceetlavarianced'uneloidePoissonsont´egalesauparam`etreλ.Eneffet

$$E(X)=\sum_{x=0}^{\infty}x\Pr(X=x)=\sum_{x=0}^{\infty}xe^{-\lambda}\frac{\lambda^x}{x!}=e^{-\lambda}\sum_{x=1}^{\infty}x\frac{\lambda^x}{x!}=e^{-\lambda}\lambda\sum_{x=1}^{\infty}\frac{\lambda^{x-1}}{(x-1)!}=e^{-\lambda}\lambda\sum_{z=0}^{\infty}\frac{\lambda^z}{z!}$$ enposant$z=x-1=e^{-\lambda}\lambda e^{\lambda}=$

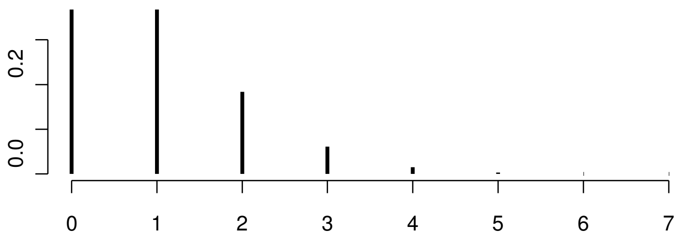


Fig.6.3–Distributiond'unevariabledePoissonavecλ=1.

```
    lwd=3,xlab="",ylab="",main="",frame=FALSE)
#PoissonP(1)plot(dpois(0:7,1),type="h",lwd=3,xlab="",ylab="",main="",frame=FALSE)
```

6.5Variableal´eatoirecontinue6.5.1D´efinition,esp´eranceetvarianceUnevar

Laprobabilit´equ'unevariableal´eatoirecontinuesoitinf´erieure`aunevaleurparticuli`ereestdonn´epar safonctionder´epartition.Pr($X\leq x$)=$F(x)$.Lafonctionder´epartitiond'unevariableal´eatoirecontinueesttoujours: ´erieure`aunevaleurquelconquevaut:Pr[$X\leq a$]=$\int_{-\infty}^{a} f(x)dx$=$F(a)$.DanslaFigure6.4,laprobabilit´ePr[$X\leq a$]estl'aire

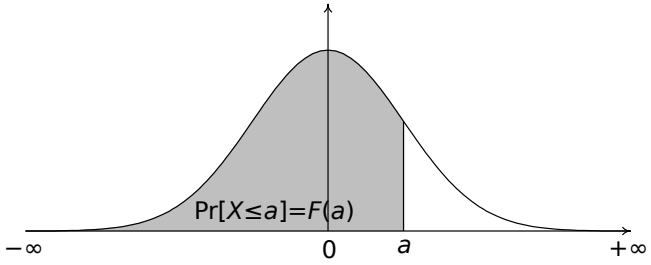


Fig.6.4–Probabilit´equelavariablealéatoiresoitinférieure`a*a*

Silavariableal´eatoireestcontinue,laprobabilit´equ'elleprenneexactementunevaleurquelconqueestnulle:

$$\Pr[X=a]=0.$$

L'esp´eranced'unevariableal´eatirecontinueestdéfinie par:

$$\mathrm{E}(X)=\int_{-\infty}^{\infty} xf(x)dx,$$

etlavariancevar$(X)=\int_{-\infty}^{\infty}(x-\mu)^2 f(x)dx.$

6.5.2Variableuniforme

Unevariableal´eatoire$X$estditeuniformedansunintervalle[a,b],(avec$a<b$)sisar´

epartitionest:

$F(x)=$□□□0si$x<a(x-a)/(b-a)$si$a\le x\le b$1si$x>b.$Sadensit´eestalors$f(x)=$□□□0si$x<a1/(b-a)$si$a\le x\le b$0si$x>b.$Onpe

✷

R´esultat6.2

$$\sigma^2 = \text{var}(X) = \frac{(b-a)^2}{12}.$$

D´emonstration Demani`ereg´en´erale,unevariancepeuttoujourss''ecrirecommeunmoment`al'originedl'ordre2moinsle carr´edelamoyenne.Eneffet,

$$\begin{aligned}
\sigma_2 &= \text{var}(X) \\
&= \int_a^b (x-\mu)^2 f(x)dx \\
&= \int_a^b (x_2+\mu^2 - 2x\mu)f(x)dx \\
&= \int_a^b x_2 f(x)dx + \int_a^b \mu^2 f(x)dx - 2\mu \int_a^b x f(x)dx \\
&= \int_a^b x_2 f(x)dx + \mu_2 - 2\mu^2
\end{aligned}$$

$= \int_a^b x_2 f(x)dx - \mu_2$.Oncalculeensuiteunmoment`al'originedl'ordre2:$\int_a^b x_2 f(x)dx = \int_a^b x_2 \frac{1}{b-a}dx = \frac{1}{b-a}\int_a^b x_2 dx = \frac{1}{b-a}$

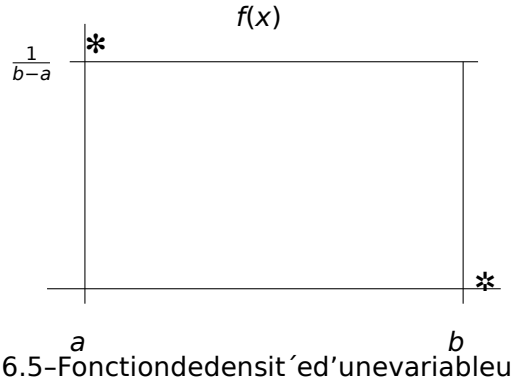


Fig.6.5–Fonctiondedensit´ed’unevariableuniforme

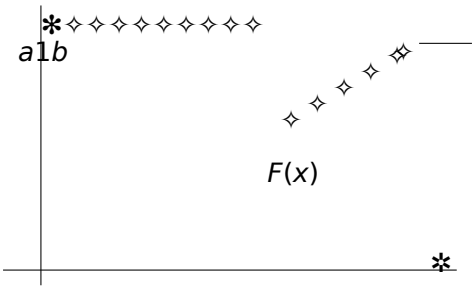
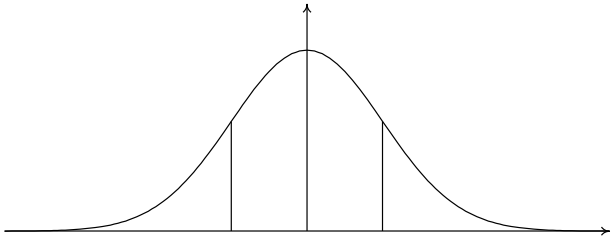


Fig.6.6–Fonctionder´epartitiond’unevariableuniforme

6.5.3Variablenormale Unevariableal´eatoire$X$estditenormalesisadensit´evaut$f_{\mu,\sigma^2}(x)=1\sigma\sqrt{2\pi}\exp-12x$

o`u$\mu\in$Ret$\sigma\in$R2sontlesparam`etresdeladistribution.Leparam`etreμestappel´elamoyenneetle
param`etreσl’´ecart-typedelavariablenormale.$\mu-\infty+\infty\mu+\sigma\mu-\sigma$Fig.6.7–Fonctiondedensit´ed’unevariablenorma



Lafonctionder´epartitionvaut

$$F_{\mu,\sigma^2}(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} \exp -\frac{1}{2}\left(\frac{u-\mu}{\sigma}\right)^2 du.$$

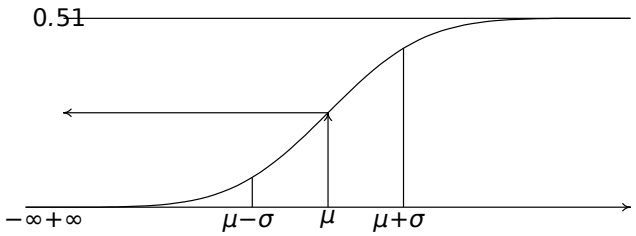
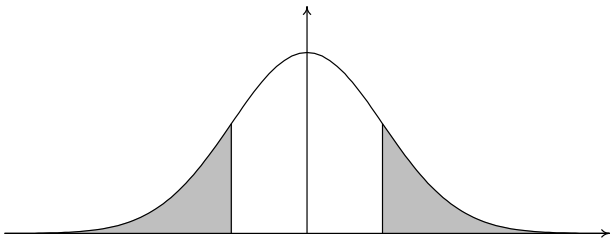


Fig.6.8–Fonctionder´ epartitiond'unevariablenormale

6.5.4Variablenormalecentr´eer´eduite Lavariableal´eatoirenormalecentr´eer´eduiteestunevari

d'esp´erancenulle$\mu=0$etde

variance$\sigma^2=1$.Safonctiondedensit´evaut$f_{0,1}(x)=1\sqrt{2\pi}\exp -x^2 2$.etsar´epartitionvaut$\Phi(x)=F_{0,1}(x)=_{x-\infty}1\sqrt{2\pi}\exp$



Enposant

$$z = \frac{u-\mu}{\sigma},$$

onobtient$u = z\sigma + \mu$,etdonc$du = \sigma dz$.Donc,

$$F_{\mu,\sigma^2}(x) = \int_{-\infty}^{\frac{x-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} \exp -\frac{z^2}{2} \sigma dz = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

✷

Lestablesdelavariablenormalenesontdonn´eesquepourlanormalecentr´eer´eduite.Lestablesne donnent$\Phi(x)$quepourlesvaleurspositivesde$x$, carlesvaleursn´egativespeuventˆetretrouv´eesparla relationdesym´etrie.

6.5.5 Distributionexponentielle

Unevariableal´eatoire$X$aunedistributionexponentiellesisafonctiondedensit´e est donn´eepar:

$$f(x) = \begin{cases} \lambda \exp -(\lambda x), & \text{si } x > 0 \\ 0 \text{ sinon} \end{cases}$$

Leparam`etreλestpositif.Quand$x > 0$,safonctionder´epartitionvaut:

$$F(x) = \int_0^x f(u)du = \int_0^x \lambda e^{-\lambda u} du = -e^{-\lambda u} \Big|_0^x = 1 - e^{-\lambda x}.$$

Onpeutalorscalculerlamoyenne:**R´esultat6.4** $E(X) = 1\lambda$ **D´emonstration** $E(X) = \int_0^\infty x f(x)dx = \int_0^\infty x\lambda e^{-\lambda x} dx = -1 + x\lambda$

$$\frac{1}{\lambda}.$$

✷

Ilest´egalementpossibledemontrerquelavariancevaut:$var(X) = 1\lambda_2$.

6.6 Distributionbivari´ee

De

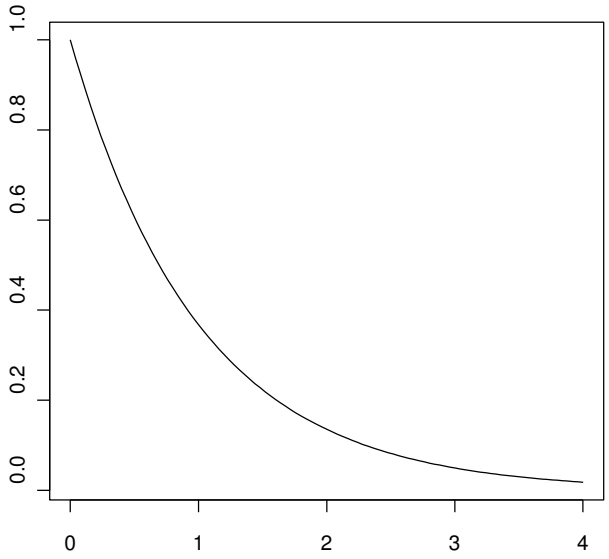


Fig.6.10–Fonctiondedensit´ed'unevariableexponentielleavec$\lambda=1$.

Aveclesdistributionsmarginales,onpeutd´efinirlesmoyennesmarginales,etlesvariancesmarginales:

$$\mu_X=\int_{-\infty}^{\infty}xf_X(x)dx,\text{et}\mu_Y=\int_{-\infty}^{\infty}yf_Y(y)dy,$$

$$\sigma^2_X=\int_{-\infty}^{\infty}(x-\mu_X)^2f_X(x)dx,\text{et}\sigma^2_Y=\int_{-\infty}^{\infty}(y-\mu_Y)^2f_Y(y)dy.$$

Onappelledensit´esconditionnelles,lesfonctions$f(x|y)=f(x,y)f_Y(y)$et$f(y|x)=f(x,y)f_X(x)$.Aveclesdistributionscond

tionnelles:$\mu_X(y)=\int_{-\infty}^{\infty}xf(x|y)dx,\text{et}\mu_Y(x)=\int_{-\infty}^{\infty}yf(y|x)dy,\sigma^2_X(y)=\int_{-\infty}^{\infty}\{x-\mu_X(y)\}^2f(x|y)dx,\text{et}\sigma^2_Y(x)=\int_{-\infty}^{\infty}\{y-\mu_Y(x)\}^2f($
´ejointedesdeuxvariables,alors$X$et$Y$sontind´ependantssi$f_{XY}(x,y)=f_X(x)f_Y(y),x,y\in$R.100

6.7 Propri´et´es des esp´erances et des variances

De mani`ere g´en´erale, pour des variables al´eatoires X et Y, et avec a et b constants, on a les r´esultats suivants. **R´esultat 6.5** $\mathrm{E}(a+bX)=a+b\mathrm{E}(X)$

D´emonstration $\mathrm{E}(a+bX)=\int_{\mathbb{R}}(a+bx)f(x)dx=a\int_{\mathbb{R}}f(x)dx+b\int_{\mathbb{R}}xf(x)dx=a+b\mathrm{E}(X).$ ✷

R´esultat 6.6 $\mathrm{E}(aY+bX)=a\mathrm{E}(Y)+b\mathrm{E}(X)$

D´emonstration $\mathrm{E}(aY+bX)=\int\int_{\mathbb{R}\mathbb{R}}(ay+bx)f(x,y)dxdy$

$$=a\int\int_{\mathbb{R}\mathbb{R}}yf(x,y)dxdy+b\int\int_{\mathbb{R}\mathbb{R}}xf(x,y)dxdy$$

$$=a\int_{\mathbb{R}}y\int_{\mathbb{R}}f(x,y)dxdy+b\int_{\mathbb{R}}x\int_{\mathbb{R}}f(x,y)dydx$$

$$=a\int_{\mathbb{R}}yf(y)dy+b\int_{\mathbb{R}}xf(x)dx=a\mathrm{E}(Y)+b\mathrm{E}(X)$$ ✷

R´esultat 6.7 $\mathrm{var}(a+bX)=b^2\mathrm{var}(X)$. **D´emonstration** $\mathrm{var}(a+bX)=\int_{\mathbb{R}}[a+bx-\mathrm{E}(a+bX)]^2f(x)dx=\int_{\mathbb{R}}[a+bx-(a+b\mathrm{E}(X$

D´emonstration

$$\begin{aligned}\mathrm{var}(X+Y) &= \int_{\mathbb{R}}\int_{\mathbb{R}} [x+y-\mathrm{E}(X+Y)]^2 f(x,y)dxdy \\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} [x-\mathrm{E}(X)+y-\mathrm{E}(Y)]^2 f(x,y)dxdy \\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} [x-\mathrm{E}(X)]^2+[y-\mathrm{E}(Y)]^2+2[x-\mathrm{E}(X)][y-\mathrm{E}(Y)]f(x,y)dxdy \\ &= \mathrm{var}(X)+\mathrm{var}(Y)+2\mathrm{cov}(X,Y)\end{aligned}$$

✷

R´esultat6.9*Deplus,siXetYsontind´ependantes,onaf*$(x,y)=f(x)f(y)$*pourtoutx,y*

$$\mathrm{E}(XY)=\mathrm{E}(X)\mathrm{E}(Y).$$

D´emonstration$\mathrm{E}(XY)=\int_{\mathbb{R}}\int_{\mathbb{R}} xyf(x)f(y)dxdy$

$$\begin{aligned}&= \int_{\mathbb{R}} xf(x)dx \int_{\mathbb{R}} yf(y)dy \\ &= \mathrm{E}(X)\mathrm{E}(Y).\end{aligned}$$

✷

Enfin,si$X$et$Y$sontind´ependantes,onacov$(X,Y)=0$,etdoncvar$(X+Y)=$var$(X)+$var(Y).Enfin,ilestpossibledecalc

ependantes,

etidentiquementdistribu´ees.**Th´eor`eme6.3***Soit*$X_1,\ldots,X_n$*unesuitedevariablesal´eatoires,ind´ependanteseti*

ees

etdontlamoyenneμ*etlavariance*σ^2*existentetsontfinies,alorssi*$\bar{X}=\frac{1}{n}\sum_{i=1}^{n} X_i$*,ona*$\mathrm{E}(\bar{X})=\mu$*,et*var$(\bar{X})=\sigma^2 n$.**D´emo**

$\bar{}$

$\bar{}$ $\bar{}$ $\bar{}$

$\bar{}$ $\bar{}$ $\bar{}$ $\bar{}$

6.8 Autres variables aléatoires

6.8.1 Variable khi-carrée

Soit une suite de variables aléatoires indépendantes, normales, centrées réduites, $X_1, \ldots, X_p$, (c'est-à-dire de moyenne nulle et de variance égale à 1), alors la variable aléatoire

$$\chi^2_p = \sum_{i=1}^{p} X_i^2,$$

est appelée variable aléatoire khi-carré à p degrés de liberté.

Il est possible de montrer que $E(\chi^2_p) = p$,

et que $\mathrm{var}(\chi^2_p) = 2p$.

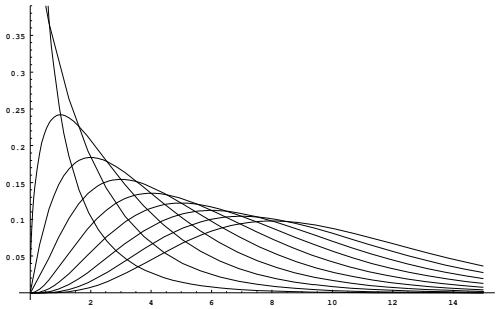


Fig. 6.11 – Densité d'une variable de khi-carré avec $p = 1, 2, \ldots, 10$

6.8.2 Variable de Student

Soit une variable

es

de liberté, indépendante de X, alors la variable aléatoire $t_p = X\chi_{2p}/p$ est appelée variable aléatoire de Student à p

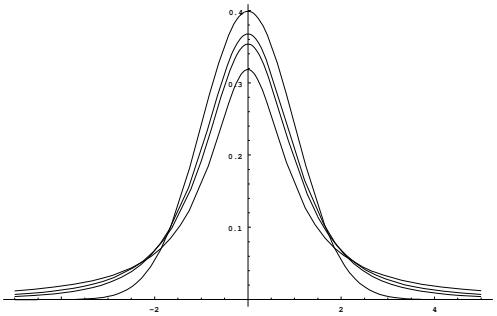


Fig. 6.12 – Densités de variables de Student avec $p = 1, 2$ et 3 et d'une variable normale

6.8.3 Variable de Fisher

Soient deux variables aléatoires khi-carré indépendantes χ^2_p, χ^2_q, respectivement à p et q degrés de liberté, alors la variable aléatoire

$$F_{p,q} = \frac{\chi^2_p/p}{\chi^2_q/q}$$

est appelée variable aléatoire de Fisher à p et q degrés de liberté.

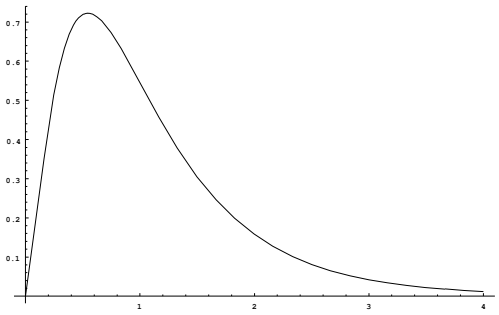


Fig. 6.13 – Densité d'une variable de Fisher

Remarque 6.1 Il est facile de montrer que le carré d'une variable de Student à q degrés de liberté est une variable de Fisher à 1 et q degrés de liberté.

6.8.4 Variable normale multivariée

Le vecteur de variables

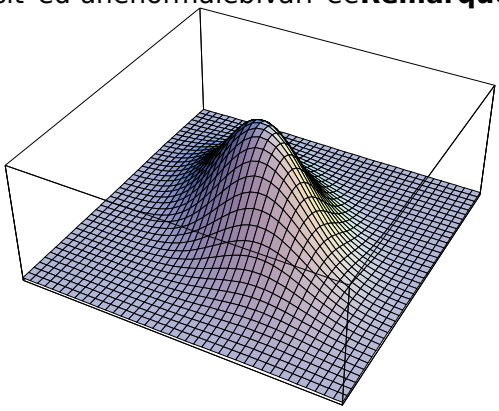
ee de moyenne $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)$ et de matrice variance-covariance $\boldsymbol{\Sigma}$ (on suppose par simplicité que $\boldsymbol{\Sigma}$ est de plein rang), si sa fonction de densité est donnée par

$$f_X(\mathbf{x}) = 1(2\pi)_{p/2}|\boldsymbol{\Sigma}|_{1/2}\exp -12(\mathbf{x}-\boldsymbol{\mu})\boldsymbol{\Sigma}_{-1}(\mathbf{x}-\boldsymbol{\mu}), \quad (6.2)$$

pour tout $\mathbf{x} \in \mathrm{R}_p$.

Fig. 6.14 – Densité d'une normale bivariée

Remarque 6.2 Si $p = 1$, on retrouve l'expression (6.1).



Un cas particulier est important : supposons que la matrice variance-covariance peut s'écrire $\mathbf{\Sigma} = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_p^2)$, ce qui signifie que toutes les composantes du vecteur $\mathbf{X}$ sont non-corrélées. Dans ce cas,

$$\begin{aligned} f_X(\mathbf{x}) &= \frac{1}{(2\pi)^{p/2}|\mathbf{\Sigma}|^{1/2}} \exp -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) \\ &= \frac{1}{(2\pi)^{p/2}(\prod_{j=1}^{p}\sigma_j^2)^{1/2}} \exp -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) \\ &= \frac{1}{(2\pi)^{p/2}(\prod_{j=1}^{p}\sigma_j)} \exp\left(-\sum_{j=1}^{p}\frac{(x_j-\mu_j)^2}{2\sigma_j^2}\right) \\ &= 1 \frac{1}{(2\pi)_{p/2}(\prod_{j=1}^{p}\sigma_j)} \prod_{j=1}^{p} \exp -\frac{(x_j-\mu_j)^2}{2\sigma_j^2} \\ &= \prod_{j=1}^{p} 1(2\pi)_{1/2}\sigma_j \exp -\frac{(x_j-\mu_j)^2}{2\sigma_j^2} \end{aligned}$$

$= \prod_{j=1}^{p} f_{X_j}(x_j)$, où $f_{X_j}(x_j) = 1(2\pi\sigma_{2j})_{1/2} \exp -(x_j-\mu$

$$\frac{}{} \qquad \frac{}{2\sigma_2} \,,$$

est la densité de la variable X_j. On constate qu'il y a absence de corrélation entre les variables normales, alors la densité du vecteur normal peut s'écrire comme un produit de densités. Dans le cas multinormal (et seulement dans ce cas), l'absence de corrélation implique donc l'indépendance des variables aléatoires.

Demani`ere g´en´erale, si $\mathbf{X}$ est un vecteur de variables al´eatoires de moyenne $\boldsymbol{\mu}$ et de matrice variance-covariance $\mathbf{\Sigma}$, et si $\mathbf{A}$ est une matrice $q \times p$ de constantes, alors $E(\mathbf{AX}) = \mathbf{A}E(\mathbf{X}) = \mathbf{A}\boldsymbol{\mu}$, et $\mathrm{var}(\mathbf{AX}) = \mathbf{A}\mathrm{var}(\mathbf{X})\mathbf{A} = \mathbf{A\Sigma A}$. Dans l

pendants a matrice variance-covariance n'est pas n´ecessairement de pleine rang). Donc, si $\mathbf{X}$ est un vecteur multinor

Exercice6.2 D´eterminezlesvaleurs$j$delavariablenormalecentr´eeréduite$Z$tellesque:

1. Pr[$Z \leq j$]=0,9332;
2. Pr[$-j \leq Z \leq j$]=0,3438;
3. Pr[$Z \leq j$]=0,0125;
4. Pr[$Z \geq j$]=0,0125;
5. Pr[$j \leq Z \leq 3$]=0,7907.

Exercice6.3 Soitunevariablealéatoire$X \sim N(53; \sigma^2 = 100)$repréSentantlerésultatd'unexamenpour un´etudiantd'unesection.D´eterminezlaprobabilitépour´queler´esultatsoitcomprisentre33,4et72,6.

Exercice6.4 Soitunevariableal´eatoire$XN(50; \sigma^2 = 100)$. Déterminezlepremierquartiledecettedistribution.**Exercice6.5** Ensupposantquelestaillesencmdes´etudiantsd'unpaysadmettentladistributionnormale

$N(172; \sigma_2 = 9)$.Ondemandededé´eterminerlepourcentageth´eorique:
a)d'´etudiantsmesurantaumoins180cm.b)d'´etudiantsdontlataillleestcompriseentre168et180.**Exercice6.6**S

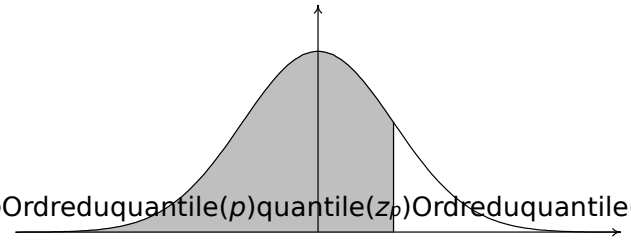
elavitesse
detouteslesautomobilespendantunejourn´ee.Ensupposantquelesvitessesrecueilliessoientdistribu´ees
normalementavecunemoyennede72km/hetun´ecart-typede8km/h,quelleestapproximativementla
proportiond'automobilesayantcommisunexc`esdevitesse?**Exercice6.7** Pourl'assemblaged'unemachine,onpr

esune
loinormaledemoyenne10cmetd'´ecart-type0,2cm.Ongroupelescylindresen3cat´egories:A:d´efectueuxetinut
´ee`a17degr´esdelibert´e;3.d'unevariabledeStudent`a8degr´esdelibert´e;4.d'unevariabledeFisher(uniquem

Chapitre 7

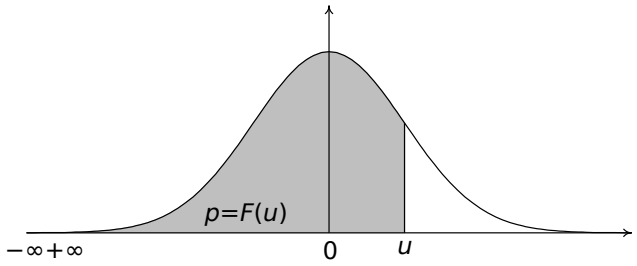
Tables statistiques

Tab. 7.1 – Table des quantiles d'une variable normale centrée réduite



$0 \quad -\infty \quad +\infty \quad z_p \quad p$ Ordre du quantile (p) quantile (z_p) Ordre du quantile (p) Quantile (z_p) 0.500 0.0000 0.975 1.96

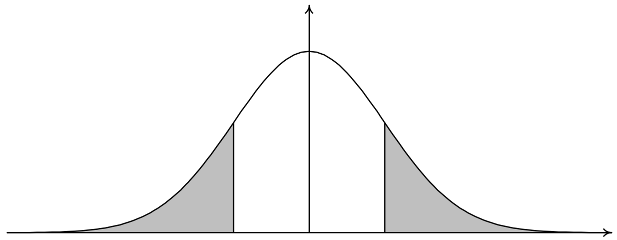
Tab.7.2 – Fonction de répartition de la loi normale centrée réduite
(*Probabilité de trouver une valeur inférieur à u*)



u	0.0	1.0	2.0	3	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599				
43	.9945	.9946	.9948	.9949	.9951	.9952				
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
7	.9997	.9998								
8										

Table 7.3 – quantiles de la loi normale centrée réduite

(*u* : valeur ayant la probabilité *α* d'être dépassée en valeur absolue)



α	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	∞	2.5758	2.3263	2.1701	2.0537	1.9600	1.880			
0.1	1.6449	1.5982	1.5548	1.5141	1.4758	1.439				
0.2	1.2816	1.2536	1.2265	1.2004	1.1750	1.150				
0.3	1.0364	1.0152	0.9945	0.9741	0.9542	0.934				

Tab.7.4–Tabledesquantilesd'unevariableχ^2`andegrésdeliberté

	ordreduquantile					
	0.01	0.025	0.05	0.95	0.975	0.99
n=1	0.000157	0.000982	0.003932	3.841	5.024	6.635
2	0.02010	0.05064	0.103	5.991	7.378	9.210
3	0.115	0.216	0.352	7.815	9.348	11.34
4	0.297	0.484	0.711	9.488	11.14	13.28
5	0.554	0.831	1.145	11.07	12.83	15.09
6	0.872	1.237	1.635	12.59	14.45	16.81
7	1.239	1.690	2.167	14.07	16.01	18.48
8	1.646	2.180	2.733	15.51	17.53	20.09
9	2.088	2.700	3.325	16.92	19.02	21.67
10	2.558	3.247	3.940	18.31	20.48	23.21
11	3.053	3.816	4.575	19.68	21.92	24.72
12	3.571	4.404	5.226	21.03	23.34	26.22
13	4.107	5.009	5.892	22.36	24.74	27.69
14	4.660	5.629	6.571	23.68	26.12	29.14
15	5.229	6.262	7.261	25.00	27.49	30.58
16	5.812	6.908	7.962	26.30	28.85	32.00
17	6.408	7.564	8.672	27.59	30.19	33.41
18	7.015	8.231	9.390	28.87	31.53	34.81
19	7.633	8.907	10.12	30.14	32.85	36.19
20	8.260	9.591	10.85	31.41	34.17	37.57
21	8.897	10.28	11.59	32.67	35.48	38.93
22	9.542	10.98	12.34	33.92	36.78	40.29
23	10.20	11.69	13.09	35.17	38.08	41.64
24	10.86	12.40	13.85	36.42	39.36	42.98
25	11.52	13.12	14.61	37.65	40.65	44.31	26	12.20	13.84	15.38	38.89	41.92	45.64	27	12.88	14.57	16.15	40
14	25.15	27.57	29.79	60.48	64.20	68.71	46	26.66	29.16	31.44	62.83	66.62	71.20	48	28.18	30.75	33.10	6

Tab.7.5–Tabledesquantilesd'unevariabledeStudent`a$n$degrésdeliberté

	ordreduquantile			
	0.95	0.975	0.99	0.995
n=1	6.314	12.71	31.82	63.66
2	2.920	4.303	6.965	9.925
3	2.353	3.182	4.541	5.841
4	2.132	2.776	3.747	4.604
5	2.015	2.571	3.365	4.032
6	1.943	2.447	3.143	3.707
7	1.895	2.365	2.998	3.499
81.860		2.306	2.896	3.355
91.833		2.262	2.821	3.250
101.8122.228			2.764	3.169
111.7962.201			2.718	3.106
121.7822.179			2.681	3.055
131.7712.1602.650				3.012
141.7612.1452.624				2.977
151.7532.1312.602				2.947
161.7462.1202.5832.921				
171.7402.1102.5672.898				
181.7342.1012.5522.878				
191.7292.0932.5392.861201.7252.0862.5282.845211.7212.0802.5182.831221.71				
.3812.648801.6641.9902.3742.639901.6621.9872.3682.6321001.6601.9842.3642				

Tab.7.6 – Tabledesquantilesd'ordre0.95d'unevariabledeFisher à n_1 et n_2 degrés de liberté

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10	12	14	16	20	30	∞
1	161.	199.	215.	224.	230.	234.	236.	238.	240.	241.	243.	245.	246.	248.	250.	254.
2	18.5	19.0	19.1	19.2	19.3	19.3	19.3	19.3	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.50
3	10.1	9.55	9.27	9.11	9.01	38.9	4	8.87	8.85	8.81	8.78	6	8.74	8.71	8.66	8.526
4	7.70	6.94	6.59	6.38	6.25	6.16	3	6.09	4	6.04	9.91	5.87	5.84	5.80	5.74	5.628
5	6.60	5.78	5.40	5.19	5.05	4.95	0	4.87	6	4.81	4.77	2	4.73	4.69	4.64	4.365
6	5.98	5.14	4.75	4.53	4.38	4.28	4	4.20	7	4.14	4.09	4	4.05	4.00	3.95	3.669
7	5.59	4.73	4.34	4.12	3.97	3.86	6	3.78	7	3.72	3.67	3	3.63	3.57	3.51	3.230
8	5.31	4.45	4.06	3.83	3.68	3.58	1	3.50	3	3.44	3.39	0	3.34	3.28	3.22	2.928
9	5.11	4.25	3.86	3.63	3.48	3.37	4	3.29	3	3.23	3.18	9	3.14	3.07	3.01	2.707
10	4.96	4.10	3.71	3.48	3.33	3.22	2	3.14	2	3.07	3.02	8	2.98	2.91	2.84	2.538
11	4.84	3.98	3.59	3.36	3.20	3.09	5	3.01	2	2.95	2.90	7	2.85	2.79	2.72	2.404
12	4.74	3.88	3.49	3.26	3.11	3.00	9	2.91	2	2.85	2.80	7	2.75	2.69	2.62	2.296
13	4.66	3.80	3.41	3.18	3.02	2.92	3	2.83	2	2.77	2.72	6	2.67	2.60	2.53	2.206
14	4.60	3.73	3.34	3.11	2.96	2.85	8	2.76	2	2.70	2.65	6	2.60	2.53	2.46	2.131
15	4.54	3.68	3.29	3.06	2.90	2.79	0	2.71	2	2.64	2.59	5	2.54	2.48	2.40	2.066
16	4.49	3.63	3.24	3.01	2.85	2.74	3	2.66	2	2.59	2.54	5	2.49	2.42	2.35	2.010
17	4.45	3.59	3.20	2.96	2.81	2.70	7	2.61	2	2.55	2.50	4	2.45	2.38	2.31	1.960
18	4.41	3.55	3.16	2.93	2.77	2.66	6	2.58	2	2.51	2.46	4	2.41	2.34	2.27	1.917
19	4.38	3.52	3.13	2.90	2.74	2.63	4	2.54	2	2.48	2.42	4	2.38	2.31	2.23	1.878
20	4.35	3.49	3.10	2.87	2.71	2.60	8	2.51	2	2.45	2.39	4	2.35	2.28	2.20	1.843
21	4.32	3.47	3.07	2.84	2.68	2.57	6	2.49	2	2.42	2.37	3	2.32	2.25	2.18	1.812
22	4.30	3.44	3.05	2.82	2.66	2.55	4	2.46	2	2.40	2.34	3	2.30	2.23	2.15	1.783
23	4.28	3.42	3.03	2.80	2.64	2.53	2	2.44	2	2.37	2.32	2	2.27	2.20	2.13	1.757
24	4.26	3.40	3.01	2.78	2.62	2.51	2	2.42	2	2.36	2.30	2	2.25	2.18	2.11	1.733
25	4.24	3.38	2.99	2.76	2.60	2.49	4	2.40	2	2.34	2.28	2	2.24	2.16	2.09	1.711
26	4.22	3.37	2.98	2.74	2.59	2.47	4	2.39	2	2.32	2.27	2	2.22	2.15	2.07	1.691
27	4.21	3.35	2.96	2.73	2.57	2.46	3	2.37	2	2.31	2.25	2	2.20	2.13	2.06	1.672
28	4.19	3.34	2.94	2.71	2.56	2.44	5	2.36	2	2.29	2.24	2	2.19	2.12	2.04	1.654
29	4.18	3.33	2.93	2.70	2.55	2.43	5	2.35	2	2.28	2.22	2	2.18	2.10	2.03	1.637
30	4.17	3.32	2.92	2.69	2.53	2.42	7	2.33	2	2.27	2.21	2	2.16	2.09	2.01	1.621
40	4.08	3.23	2.84	2.61	2.45	2.34	5	2.25	2	2.18	2.12	2	2.08	2.00	1.92	1.509
50	4.03	3.18	2.79	2.56	2.40	2.29	0	2.20	2	2.13	2.07	2	2.03	1.95	1.87	1.432
60	4.00	3.15	2.76	2.53	2.37	2.25	8	2.17	2	2.10	2.04	1	1.99	1.92	1.84	1.389
70	3.98	3.13	2.74	2.50	2.35	2.23	7	2.14	2	2.07	2.02	1	1.97	1.89	1.81	1.360
80	3.96	3.11	2.72	2.49	2.33	2.21	6	2.13	2	2.06	2.00	1	1.95	1.88	1.79	1.340
90	3.95	3.10	2.71	2.47	2.32	2.20	5	2.11	2	2.04	1.99	1	1.94	1.86	1.78	1.324
100	3.94	3.09	2.70	2.46	2.31	2.19	4	2.10	2	2.03	1.97	1	1.93	1.85	1.77	1.311
120	3.92	3.07	2.68	2.45	2.29	2.18	2	2.09	2	2.01	1.96	1	1.91	1.83	1.75	1.290
140	3.91	3.06	2.67	2.44	2.28	2.17	2	2.08	2	2.00	1.95	1	1.90	1.82	1.74	1.274
160	3.90	3.05	2.66	2.43	2.27	2.16	2	2.07	2	1.99	1.94	1	1.89	1.81	1.72	1.261
180	3.89	3.04	2.65	2.42	2.27	2.15	4	2.06	2	1.98	1.93	1	1.88	1.80	1.72	1.250
200	3.89	3.04	2.65	2.42	2.26	2.14	3	2.06	2	1.98	1.93	1	1.87	1.79	1.71	1.240
250	3.88	3.03	2.64	2.41	2.25	2.14	1	2.05	2	1.97	1.92	1	1.87	1.79	1.70	1.223
300	3.87	3.03	2.64	2.41	2.25	2.13	9	2.04	2	1.97	1.91	1	1.86	1.78	1.70	1.211
400	3.86	3.02	2.63	2.40	2.24	2.13	3	2.03	2	1.96	1.91	1	1.85	1.78	1.69	1.202
500	3.86	3.01	2.62	2.40	2.24	2.12	9	2.03	2	1.96	1.90	1	1.85	1.77	1.69	1.194
1000	3.85	3.00	2.61	2.39	2.23	2.12	2	2.02	2	1.95	1.90	1	1.84	1.76	1.68	1.185
∞	3.84	2.99	2.60	2.37	2.21	2.09	9	2.01	1	1.94	1.88	1	1.83	1.75	1.67	1.167

Tab. 7.7 – Table des quantiles d'ordre 0.99 d'une variable de Fisher à n_1 et n_2 degrés de liberté

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10	12	14	16	20	30	∞
1	4052	5000	5403	5625	5764	5859	5928	5981	6022	6056	6106	6143	6170	6209	6261	6366
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.44	99.45	99.47	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.92	26.83	26.69	26.51	26.13
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.25	14.15	14.02	13.84	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.888	9.770	9.680	9.553	9.379	9.020
6	13.75	10.92	9.780	9.148	8.746	8.466	8.260	8.102	7.976	7.874	7.718	7.605	7.519	7.396	7.229	6.880
7	12.25	9.547	8.451	7.847	7.460	7.191	6.993	6.840	6.719	6.620	6.469	6.359	6.275	6.155	5.992	5.650
8	11.26	8.649	7.591	7.006	6.632	6.371	6.178	6.029	5.911	5.814	5.667	5.559	5.477	5.359	5.198	4.859
9	10.56	8.022	6.992	6.422	6.057	5.802	5.613	5.467	5.351	5.257	5.111	5.005	4.924	4.808	4.649	4.311
10	10.04	7.559	6.552	5.994	5.636	5.386	5.200	5.057	4.942	4.849	4.706	4.601	4.521	4.405	4.247	3.909
11	9.646	7.206	6.217	5.668	5.316	5.069	4.886	4.744	4.632	4.539	4.397	4.293	4.213	4.099	3.941	3.602
12	9.330	6.927	5.953	5.412	5.064	4.821	4.640	4.499	4.388	4.296	4.155	4.052	3.972	3.858	3.701	3.361
13	9.074	6.701	5.739	5.205	4.862	4.620	4.441	4.302	4.191	4.100	3.960	3.857	3.778	3.665	3.507	3.165
14	8.862	6.515	5.564	5.035	4.695	4.456	4.278	4.140	4.030	3.939	3.800	3.698	3.619	3.505	3.348	3.004
15	8.683	6.359	5.417	4.893	4.556	4.318	4.142	4.004	3.895	3.805	3.666	3.564	3.485	3.372	3.214	2.868
16	8.531	6.226	5.292	4.773	4.437	4.202	4.026	3.890	3.780	3.691	3.553	3.451	3.372	3.259	3.101	2.753
17	8.400	6.112	5.185	4.669	4.336	4.102	3.927	3.791	3.682	3.593	3.455	3.353	3.275	3.162	3.003	2.653
18	8.285	6.013	5.092	4.579	4.248	4.015	3.841	3.705	3.597	3.508	3.371	3.269	3.190	3.077	2.919	2.566
19	8.185	5.926	5.010	4.500	4.171	3.939	3.765	3.631	3.523	3.434	3.297	3.195	3.116	3.003	2.844	2.489
20	8.096	5.849	4.938	4.431	4.103	3.871	3.699	3.564	3.457	3.368	3.231	3.130	3.051	2.938	2.778	2.421
21	8.017	5.780	4.874	4.369	4.042	3.812	3.640	3.506	3.398	3.310	3.173	3.071	2.993	2.880	2.720	2.360
22	7.945	5.719	4.817	4.313	3.988	3.758	3.587	3.453	3.346	3.258	3.121	3.019	2.941	2.827	2.667	2.305
23	7.881	5.664	4.765	4.264	3.939	3.710	3.539	3.406	3.299	3.211	3.074	2.973	2.894	2.781	2.620	2.256
24	7.823	5.614	4.718	4.218	3.895	3.667	3.496	3.363	3.256	3.168	3.032	2.930	2.852	2.738	2.577	2.211
25	7.770	5.568	4.675	4.177	3.855	3.627	3.457	3.324	3.217	3.129	2.993	2.892	2.813	2.699	2.538	2.169
26	7.721	5.526	4.637	4.140	3.818	3.591	3.421	3.288	3.182	3.094	2.958	2.857	2.778	2.664	2.503	2.131
27	7.677	5.488	4.601	4.106	3.785	3.558	3.388	3.256	3.149	3.062	2.926	2.824	2.746	2.632	2.470	2.097
28	7.636	5.453	4.568	4.074	3.754	3.528	3.358	3.226	3.120	3.032	2.896	2.795	2.716	2.602	2.440	2.064
29	7.598	5.420	4.538	4.045	3.725	3.499	3.330	3.198	3.092	3.005	2.868	2.767	2.689	2.574	2.412	2.034
30	7.562	5.390	4.510	4.018	3.699	3.473	3.304	3.173	3.067	2.979	2.843	2.742	2.663	2.549	2.386	2.006
40	7.314	5.179	4.313	3.828	3.514	3.291	3.124	2.993	2.888	2.801	2.665	2.563	2.484	2.369	2.203	1.805
50	7.171	5.057	4.199	3.720	3.408	3.186	3.020	2.890	2.785	2.698	2.562	2.460	2.380	2.265	2.098	1.715
60	7.077	4.977	4.126	3.649	3.339	3.119	2.953	2.823	2.718	2.632	2.496	2.394	2.314	2.198	2.030	1.640
80	6.963	4.881	4.036	3.563	3.255	3.036	2.871	2.742	2.637	2.551	2.415	2.312	2.232	2.115	1.946	1.549
100	6.895	4.824	3.984	3.513	3.206	2.988	2.823	2.694	2.590	2.503	2.367	2.265	2.184	2.067	1.896	1.493
120	6.851	4.787	3.949	3.480	3.174	2.956	2.792	2.663	2.559	2.472	2.336	2.233	2.152	2.035	1.863	1.455
200	6.763	4.713	3.881	3.414	3.111	2.894	2.730	2.601	2.497	2.411	2.274	2.171	2.090	1.971	1.797	1.381
500	6.686	4.648	3.821	3.357	3.054	2.838	2.674	2.545	2.442	2.355	2.218	2.114	2.033	1.913	1.735	1.306
∞	6.635	4.605	3.782	3.319	3.017	2.802	2.639	2.511	2.407	2.321	2.183	2.079	1.997	1.877	1.695	1.245

Listedestableaux

Tabledesfigures

Index